Handle with care

The possibility that H7N9 avian influenza may evolve sufficiently to cause a pandemic has scientists turning again to controversial research — they must be careful how they justify the risks taken.

The H7N9 avian flu virus first reported in China in March has so far infected at least 134 people, and killed 43 of them. Thankfully, there are no signs yet that it can easily be transmitted between people — instead it is sporadically being caught by humans through contact with chickens and other fowl.

Researchers now want to make genetically engineered versions of H7N9 that are more transmissible and pathogenic in mammals. In a Correspondence published jointly this week in *Nature* and *Science* (see page 150), 22 scientists, including Ron Fouchier of the Erasmus Medical Center in Rotterdam, the Netherlands, and Yoshihiro Kawaoka of the University of Wisconsin-Madison, argue that such research can help to assess the 'pandemic potential' of H7N9. The dilemma is that should such engineered strains be accidentally or deliberately released from a lab, they could spark a flu pandemic.

The announcement is likely to prompt some replay of last year's debate over the creation by Fouchier and Kawaoka of lab strains of H5N1 that could transmit between ferrets. And it offers the first test of some of the review and oversight structures put in place for this 'gain-of-function' flu research. As this journal has said before, scientists who push for such research should be wary of over-selling the benefits to public health, at least in the short term, as a way to justify the risks taken.

A sense of perspective is crucial here. The long-term benefits of such work are clear — as long as it is done to the highest biosafety standards. It will shed light on, for example, the mechanisms of virus transmissibility and pathogenicity. But the immediate benefits to public health and our short-term ability to counter the threat of H7N9 are less clear-cut. Scientists cannot predict pandemics, so to assess the pandemic potential of viruses — and to decide which strains warrant the manufacture

of trial vaccines — comes down to judgements of relative risk.

Tests of how flu viruses behave in animal models such as ferrets can certainly provide information on the risk of transmissibility and pathogenicity, although it can be difficult to extrapolate those results to humans. A rash of papers this year has shown that H7N9 does have limited airborne transmissibility in ferrets, although the virus is not transmitting between people in the current outbreak in China.

Another way to assess pandemic potential is to monitor wild-type viruses for mutations that allow the virus more ready access to human cells. H7N9 has already acquired some of these mutations, which is why it infects humans more easily than does H5N1. But as researchers pointed out in June, there is no scientific evidence that such mutations predict the risk of a pandemic (D. M. Morens *et al. N. Engl. J. Med.* **368,** 2345–2348; 2013). Transmissibility is more complex than that.

In creating mammalian- transmissible versions of H7N9, scientists would go a step further and hope to identify combinations of mutations that could increase virus transmissibility in ferrets or other models. Such work could yield information on the biological principles affecting transmission. But nature could well come up with combinations for transmission that are different from those obtained in experiments.

Following the H5N1 controversy, the US Department of Health and Human Services has introduced an extra layer of review that will apply to anyone seeking funding for work to make mammalian-transmissible strains of H7N9 (see page 151). The risks and benefits of the work will be assessed by a panel of experts in public health, security, risk assessment, law and ethics, and, importantly, any extra steps needed to mitigate biosafety risks will be considered. The way the review handles H7N9 will be an important test of the effectiveness and transparency of this new approach. ∎

Blood ties

Scientists should give donors more information about how their biospecimens are used.

The family of Henrietta Lacks is finally getting a say in how researchers can use her cells, six decades after her fatal cervical tumour spawned the HeLa cell line. There is little doubt that the controversy over the case contributed to the decision by the US National Institutes of Health to consult her relatives about the future use of her genome information (see pages 132 and 141). But people who donate samples to biomedical research today are unlikely to find out what happens to their material.

Standards of informed consent have improved since scientists

established HeLa without approval from Lacks or her family. But research participants still have little control over how their tissues and data are used, and often never hear from the researchers again.

Increasingly, volunteers are asked to give 'broad consent' for samples and data to be used in studies that may not have been conceived at the time of donation. In exchange, donors should have the option to learn how their specimens are being used — and even to withdraw consent.

This already happens informally in some studies, but digital technologies could allow researchers to keep patients updated. Imagine the thrill of giving a sample, logging on to a secure website years later and discovering that your specimen helped to develop a skin-cancer treatment.

This continued contact with donors raises issues — not least how to ensure their anonymity. But researchers must also be honest and tell donors that privacy cannot be guaranteed, particularly for highly identifiable genomic information. Some volunteers and their families are rightly proud that they are directly contributing to research. Funders and researchers should give more of them the chance to stay involved. ∎

Balancing privacy with public benefit

Maximizing access to research data will greatly benefit science, and users can help to establish universal principles on how to do it, says **Martin Bobrow**.

This week's publication in _Nature_ of a second HeLa cancer-cell genome (see page 207) and the announcement from the US National Institutes of Health on how it will control who gets to use the sequence information (see page 141) highlight a growing issue in modern science: access to biomedical and health-related research data.

The amount of such data continues to grow at breakneck speed, generated by large epidemiological and cohort studies that track people's health over many years (for example, the UK Biobank project) and by studies that sequence the DNA of many individuals, such as the 1000 Genomes project. Researchers, funders and governments are becoming increasingly aware of the potential power of linking and co-analysing different data sets. Genomic data linked to large sets of patient records, for example, might reveal connections about disease that we would not otherwise discover, and data from the social sciences could add further value to these studies.

Maximizing access to data resources should increase the chances that scientists will make discoveries with medical benefits. As a result, most major research funders require grant recipients to make any large data sets they create available to other researchers. It is an ethical imperative that we seek to maximize the value of research data generated from human participants, particularly when using public funds.

In response to open-access policies, a trend is emerging of allowing legitimate researchers access to research data before publication. In making unpublished data available, however, two sets of interests need to be safeguarded. Most research participants expect privacy protection and do not want their genomes or health records to be readily identifiable. Furthermore, researchers who spend time, effort and ingenuity to generate, process and manage large research data sets expect to get appropriate credit. This also relates to emerging discussions about clinical trials: there is a need for more access to patient-level data (as highlighted by the AllTrials campaign), while respecting the terms of study participants' consent.

To navigate these issues, many large genome and longitudinal studies have set up specific data-access procedures, often overseen by committees. This is what the National Institutes of Health has done for the HeLa sequence. As the number of these data-access committees grows along with the links between data sets, a question arises: is such a piecemeal approach appropriate? The scientific and medical potential of data will only be realized if researchers are not stymied by myriad data-access mechanisms and by inconsistent ways of recording and describing data variables. So, does biomedical science need to establish and enforce common principles of governance?

There are reasons to be cautious: linking data is likely to increase the risk of individuals being identified. As more research based on linked data sets emerges, it will be extremely important to understand how these data are being used, to quantify the risks and to devise proportionate governance that allows innovative uses of data to flourish while protecting participant confidentiality as far as possible.

I chair the Expert Advisory Group on Data Access — a working group that has been set up to provide strategic advice on this issue to funders — and we need your help. We have already talked to those who produce and manage biomedical and social-science data. Now we want to hear from those who use the data, or who would like to use them in future.

What does the regulatory landscape look like for potential data users? Are we maximizing the value of the data, and if not, why not? How many data sets are out there, in what fields, governed by how many data-access committees, operating to what standards?

> **IT IS AN ETHICAL IMPERATIVE THAT WE SEEK TO MAXIMIZE THE VALUE OF RESEARCH DATA.**

The remit of the working group is for UK-based funders, but our scope is international and we want to reach across both disciplinary and national borders. Still, so far we have found it extremely difficult to get an overview of the situation in the United Kingdom, let alone internationally. This is partly because of the proliferation of data in increasingly large, complex and heterogeneous data sets, but also because of the patchwork of regulations, standards and policies that govern the management of research data across the world.

To help fill in the gaps, we are conducting an online survey of users of research data, and we would value the input of _Nature_ readers. If you use shared data in your research, wherever you are in the world, I urge you to participate (see go.nature.com/bmun1x).

We are interested in, for example, how controlled access to data affects your projects and how easy you find it to locate the right data sets. We know that some types of data are held in central repositories, but others are held, managed and formatted within local or institutional data-management systems that are known only to a small group of collaborators.

The challenge in data access is to achieve an appropriate balance. On the one hand, managers need to rigorously safeguard the interests of research participants and to apply serious sanctions against anyone who wilfully misuses their data. On the other hand, they also need to ensure that research data are accessible to legitimate researchers without undue costs and delay. All of this will be greatly helped if there is wide agreement on principles for structure, governance and use of shared data. ■

↻ **NATURE.COM**
Discuss this article online at:
go.nature.com/gp5hgu

Martin Bobrow is an emeritus fellow at the University of Cambridge, UK.
e-mail: mb238@cam.ac.uk

RESEARCH HIGHLIGHTS

Selections from the scientific literature

Flu makes bacteria go bad

A normally harmless bacterial biofilm can take a dangerous turn when exposed to a virus.

Streptococcus pneumoniae can trigger bacterial pneumonia, but also colonizes the nose or throat passages of up to 15% of healthy adults. To learn how it might become pathogenic, researchers led by Anders Hakansson at the State University of New York in Buffalo grew films of *S. pneumoniae* on a layer of human epithelial cells of the type that normally lines airways.

Infecting the epithelial cells with influenza virus caused bacteria to disperse from the biofilms, as did mimicking flu-induced conditions such as fever. Dispersed *S. pneumoniae* invaded normally uncolonized sites including the lung and middle ear in mouse studies, where they caused oedema and inflammation. They also had a more virulent gene-expression pattern than bacteria growing in biofilms or in standard laboratory conditions. 'Interkingdom signalling' could be key to inducing disease, the authors say.
mBio **4**, e00438-13 (2013)

Heavy-metal stars make lead clouds

Two helium-rich stars contain more of the element lead than astronomers have ever seen. The stars may represent an intermediate stage of stellar evolution in which heavy metals can become enriched and form cloud-like layers.

A team led by Naslim Neelamkodan and Simon Jeffery of Armagh Observatory near Belfast,

UK, analysed light from nine subdwarf stars and discovered the two lead-rich stars about 250 parsecs and 300 parsecs from Earth, respectively. The researchers think that the stars could each contain as much as 100 billion tonnes of lead — up to 100 times the amount found in normal subdwarf stars.

Starlight exerts radiative pressure, a slight force that can nudge particles. This may drive the heavy metal to separate into a thin atmospheric layer, the team suggests.
Mon. Not. R. Astron. Soc. http://dx.doi.org/10.1093/mnras/stt1091 (2013)

FRANS LANTING/FLPA

Travelling zebras forecast the weather

Zebras in Botswana heed subtle weather and vegetation clues when choosing when and how to move to greener pastures.

Hattie Bartlam-Brooks of the University of Bristol, UK, and her colleagues fitted adult zebras (**pictured**) with tracking collars and monitored them daily on their annual migration from the Okavango Delta to the Makgadikgadi grasslands, around 250 kilometres away. They compared the migrations of seven mares with models informed by satellite data on regional vegetation and rainfall. The animals seemed to use local cues to anticipate the food and water available at their destination and adjust their movements accordingly — for example, by delaying departure or reversing direction when rainfall was unseasonably late.

Zebras' forecasting skills might help them to adjust to environmental and climate change, the authors note.
J. Geophys. Res. Biogeo. http://dx.doi.org/10.1002/jgrg.20096 (2013)

Molecular switches in RNA

An RNA–protein complex regulates gene expression in an unanticipated way.

The cellular machines called spliceosomes reconfigure transcribed RNA into its mature, protein-coding form. The 'minor spliceosome' is less than 1% as abundant as the major spliceosome, but exists in plants, fungi and animals, with precursors to hundreds of human messenger RNAs containing a section removed only by the minor spliceosome. Researchers led by Gideon Dreyfuss at the University of Pennsylvania School of Medicine in Philadelphia showed that a component of the minor spliceosome, an RNA–protein complex called U6atac, is extremely unstable and normally limits the machine's activity. When cells are stressed, however, signalling enzymes stabilize U6atac, boosting its levels and increasing the production of mature RNAs. This allows the minor

spliceosome to perform as a quick-acting valve to switch on the production of certain proteins.
eLife **2**, e00780 (2013)

SOCIAL BIOLOGY

Insects show savvy mob mentality

Among social insects, individuals excel at easy decisions and colonies do better at subtle distinctions — but big swarms may snag the best options.

When house hunting, *Temnothorax* ants (**pictured**) recruit each other to potential new homes. Takao Sasaki at Arizona State University in Tempe and his colleagues forced isolated ants and whole colonies to choose between two nest sites. When one site was only slightly better than another, the colonies did better than individuals at picking the best site. When site differences were larger, the ants performed better as individuals.

But big colonies are better when there are more choices to be made. Timothy Schaerf and colleagues at the University of Sydney, Australia, used mathematical models and field experiments to compare how honeybees (*Apis mellifera*) picked out a new nesting site. Swarms of 15,000 bees chose ideal locations more often than those with 5,000 bees. Huge swarms contain more scouts, and so can collect and compare more options for potential digs.
J. R. Soc. Interface **10**, 20130533 (2013); *Proc. Natl Acad. Sci. USA* http://dx.doi.org/10.1073/pnas.1304917110 (2013)

EVOLUTION

Mammals and monogamy

Some mammals may have turned to pair-living because of infanticide or isolated females.

Using an evolutionary tree of 230 primates as a framework, Christopher Opie of University College London and his colleagues ran simulations of evolutionary history to investigate what conditions might produce the behaviours of modern primates. They conclude that monogamy arose after males began guarding females to stop rivals from killing their offspring.

Tim Clutton-Brock and Dieter Lukas at the University of Cambridge, UK, used a similar method to study how monogamy came about in mammals generally. Using an evolutionary tree of more than 2,000 species, they found that monogamy tended to arise when females lived alone and were widely dispersed. Pair-living probably arose because males could not cover a large enough area to monopolize more than one female.
Proc. Natl Acad. Sci. USA http://dx.doi.org/10.1073/pnas.1307903110 (2013); *Science* **341**, 526–530 (2013)
For a longer story on this research, see go.nature.com/glatpz

MOLECULAR PSYCHIATRY

A factor for autism and schizophrenia

Deficits in a protein that binds RNA may be a common risk factor for disorders including schizophrenia, autism and cognitive impairment.

Nelson Freimer at the University of California, Los Angeles, Utz Fischer of the University of Würzburg, Germany, and their colleagues studied a population in northern Finland in which such disorders are particularly frequent. They discovered that many people in this region are missing a small part of a chromosome that includes the gene for the RNA-binding protein TOP3β. The mutation increases the risk of schizophrenia and several other neurodevelopmental disorders.

TOP3β interacts with FMRP, a protein associated with Fragile X syndrome and autism. In a separate paper, Sige Zou and Weidong Wang of the National Institute on Aging in Baltimore, Maryland, and their colleagues characterized these interactions and show that mutations in TOP3β or in FMRP can cause abnormal development of synapses in flies and mice.
Nature Neurosci. http://dx.doi.org/10.1038/nn.3484; http://dx.doi.org/10.1038/nn.3479 (2013)

PLANT BIOLOGY

Cheating orchids turn over new leaf

Many orchids lure pollinators with unfulfilled promises of nectar, but some former deceivers have evolved the ability to produce the sugary reward.

Up to 40% of orchid species are thought to be floral fibbers, including many members of the diverse *Disa* genus of African orchids (*Disa uniflora* **pictured**). Steven Johnson and his colleagues at the University of KwaZulu-Natal in South Africa analysed 111 *Disa* species, characterizing each for the presence or absence of nectar.

By mapping these data on to an evolutionary tree of the genus, the researchers showed that nectar production evolved from deceitful ancestors nine times and was lost once. The authors speculate that such transitions may be guided by ecological circumstances that favour the fitness of one system over the other.
Biol. Lett. **9**, 20130500 (2013)

↻ **NATURE.COM**
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

SEVEN DAYS

The news in brief

Legalizing pot

Uruguay has moved closer to becoming the first country to legalize and regulate marijuana at a national level, after the country's House of Representatives approved a controversial bill on 31 July. Backed by President José Mujica and his Broad Front coalition, the measure passed by 50 votes to 46. The bill is also expected to pass the Senate. The United Nations' International Narcotics Control Board criticized the move, saying that such a law, if enacted, would violate international drug-control treaties to which Uruguay is a party.

Redefine cancer

The word 'cancer' should be used to define only tumours or lesions that are likely to become lethal if left untreated, recommends a working group of the US National Cancer Institute. In an editorial published in the *Journal of the American Medical Association* on 29 July, the authors argue that the use of the term to describe non-life-threatening conditions can lead to needless treatment (L. J. Esserman *et al. J. Am. Med. Assoc.* http://doi.org/nb9; 2013). The group suggests that physicians should reduce screening frequency and focus on high-risk patients. It also calls for the development of better methods to tell aggressive conditions from non-threatening ones.

UC open access

The University of California (UC) faculty has adopted an open-access policy for research articles authored by its members. The policy was adopted on 24 July and publicly announced on 2 August, and it will be phased in over the

Double coffin found at Grey Friars

Archaeologists at the University of Leicester, UK, last week announced a second major find at the site of Grey Friars Church — a limestone coffin that, when opened, revealed an inner coffin of lead (pictured). A pair of feet visible through a hole in the otherwise mostly intact lead casket probably belongs to one of three prestigious figures known to be buried at the site, the researchers say. The team unearthed the 2-metre-long stone coffin last September, but was not able to open it until last month. The lead coffin will be opened after researchers can determine the approach least likely to damage its contents. In September 2012, the team discovered the remains of English king Richard III at the site.

next year. The ten-campus UC system is the latest of more than 175 universities to make research freely available. But, as with other institutions, researchers can choose to opt out of the policy — a provision that critics say renders it toothless. See go.nature.com/xydons for more.

Animal restrictions

The Italian parliament on 31 July agreed on extreme restrictions for animal research. The restrictions, which include banning the use of animals in addiction studies, were added as amendments during the implementation of a 2010 European Union (EU) directive that is already considered to impose some of the strictest regulations on animal research in the world. Critics note that individual EU member states cannot legally add further amendments to the directive, and the law has yet to be finalized by the Italian government. See go.nature.com/nmtb4x for more.

MIT Swartz report

A report commissioned by the Massachusetts Institute of Technology (MIT) in Cambridge highlights the university's failure to take leadership in the case of Aaron Swartz, a programmer and Internet activist who committed suicide in January (see go.nature.com/5wmeld). Swartz, 26, was awaiting trial for illegally downloading millions of articles from the JSTOR archive through the university's network. The review, released on 30 July and led by MIT computer scientist Hal Abelson, noted that the university took a 'hands-off' approach to the case and did not wield its expertise and reputation to shape information-policy reform.

Fraud case settled

Northwestern University in Evanston, Illinois, will pay the US government US$2.93 million to settle claims that cancer researcher Charles Bennett misused federal research grants from

the National Institutes of Health to pay for family travel and to employ unqualified relatives as consultants between 2003 and 2010. Northwestern did not admit liability in the settlement, announced on 30 July. Bennett, who now directs the Center for Medication Safety and Efficacy at the South Carolina College of Pharmacy in Columbia, has denied the allegations through his lawyer.

Beefed-up burger

Researchers have served up the first order of lab-grown beef burgers, pan-fried for a 5 August press event in London. Mark Post, a tissue engineer at Maastricht University in the Netherlands, and his colleagues created the synthetic patties from cow muscle cells, which form strands when grown in the lab. About 20,000 strands make up a fatless patty, which costs more than €250,000 (US$332,000) to make. The event's two tasters, a writer and a researcher, said that the burgers were beefy but not juicy.

Science writer dies

Science writer David Dickson has died, it was reported on 2 August. Born in 1947, Dickson worked as *Nature*'s news correspondent in Washington DC from 1978 to 1982. He also worked for *Science* and *New Scientist*, and later returned to *Nature* as news editor (**pictured** in the *Nature* office in the mid-1990s). In 2001, Dickson founded SciDev.Net, a science news service for the developing world, and served as the organization's director and editor. He retired in 2012. The Association of British Science Writers honoured Dickson last year with a lifetime achievement award. Philip Campbell, *Nature*'s editor-in-chief, paid tribute to David's "powerful combination of deep knowledge, uncompromising standards and relentless advocacy".

NSF nominee

On 31 July, astrophysicist France Córdova was announced as US President Barack Obama's choice to lead the National Science Foundation (NSF). If confirmed by the Senate, she would replace Subra Suresh, who left the US$7-billion agency in March. Córdova, a former chief scientist for NASA, was president of Purdue University in West Lafayette, Indiana, from 2007 to 2012. See go.nature.com/ufpoqx for more.

NOAA head

Oceanographer and former astronaut Kathryn Sullivan was nominated on 1 August to lead the US National Oceanic and Atmospheric Administration (NOAA). Sullivan, who in 1984 became the first US woman to walk in space, has been the acting administrator at the US$5-billion agency since February. She now awaits Senate confirmation. See go.nature.com/e371zk for more.

Antibiotics deal

Cubist Pharmaceuticals announced on 30 July that it would spend US$1.6 billion to expand its antibiotics pipeline by buying two companies, Trius Therapeutics, based in San Diego, California, and Optimer Pharmaceuticals of Jersey City, New Jersey. The deal will give Cubist, headquartered in Lexington, Massachusetts, a number of early-stage antibiotic drug candidates, as well as Dificid (fidaxomicin), Optimer's marketed treatment for *Clostridium difficile*-associated diarrhoea, and tedizolid, Trius's antibiotic against drug-resistant *Staphylococcus aureus*, which is in late-stage clinical trials.

11–16 AUGUST
The Gordon Research Conference on Biology of Aging takes place in Lucca, Italy, highlighting developments in knowledge about the loss of molecular, cellular and organismal stability during ageing.
go.nature.com/nsv8c4

14–16 AUGUST
Computer scientists and information experts discuss advances in computer, mobile and network security at the 22nd USENIX Security Symposium in Washington DC.
go.nature.com/lqaya8

Grants cancelled

The US National Science Foundation (NSF) has scrapped political-science funding opportunities for the remainder of 2013. The agency, which normally spends about US$10 million annually on such research, has not stated its reasons for the decision. But political scientists speculate that it is related to a congressional restriction, signed into law by US President Barack Obama on 26 March, that requires NSF-funded political-science research to benefit either national security or economic interests. See go.nature.com/jmligd for more.
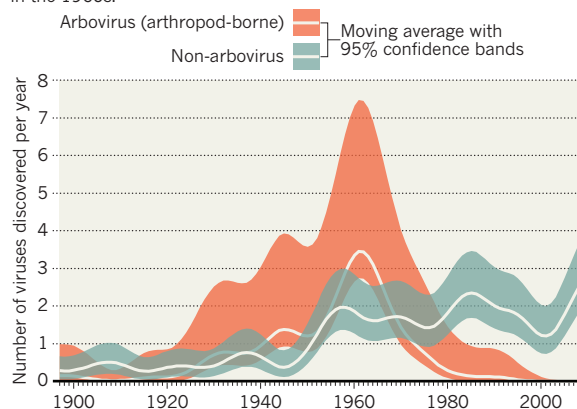
⮑ **NATURE.COM**
For daily news updates see:
www.nature.com/news

TREND WATCH

An analysis has revealed the importance of a large programme to find arthropod-borne viruses (arboviruses), sponsored by the Rockefeller Foundation in New York. Other pathogenic viruses have been discovered at a constant rate of about 2 per year for 60 years; but the 1951–66 Rockefeller programme coincided with a surge in arbovirus discovery, say researchers (R. Rosenberg *et al. Proc. Natl Acad. Sci. USA*, http://dx.doi.org/10.1073/pnas.1307243110; 2013).

THE PACE OF HUMAN VIRUS DISCOVERY

The discovery of arthropod-borne viruses has declined since a peak in the 1960s.

Arbovirus (arthropod-borne)
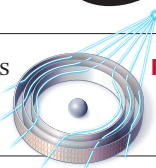Non-arbovirus
Moving average with 95% confidence bands

Number of viruses discovered per year (y-axis: 0 to 8)
(x-axis: 1900, 1920, 1940, 1960, 1980, 2000)

NEWS IN FOCUS

Chemical engineer Kemal Gürüz was taken to court in Turkey as part of controversial terrorism trials.

POLITICS

Scientists swept up in terrorism trials

Turkish government ignores calls that trials are unfair.

BY ALISON ABBOTT

"It was so emotional — we were crying after we left him," says Hans-Peter Zenner of his visit to academic Fatih Hilmioğlu in the Silivri prison, some 80 kilometres from Istanbul.

Zenner, a physician at the University of Tubingen in Germany, travelled to the penal facility in February as part of a small delegation to investigate cases of Turkish academics charged with terrorism offences. The delegation had been commissioned on behalf of an international human-rights network representing academies and scholarly societies including the US National Academies of Science and the German National Academy of Sciences Leopoldina. It concluded in its draft report on 1 August that the scientists had not received fair trials.

On 5 August, the defendants — all former or current university rectors — received harsh sentences. Hilmioğlu, a physician and former rector of İnönü University in Malatya, was sentenced to 23 years in prison on charges of conspiring to destabilize the government through political violence. Four other academics — including a transplant surgeon and chemical engineer — were given sentences of between 10 and 15 years. Another, who had been in detention for more than 4 years, was released despite being sentenced to 12 years and 6 months in prison.

The delegation had concluded in its report that "standards of justice failed", and that in no case did the evidence brought by prosecutors "support the conclusion that any of our … colleagues is guilty of committing the crimes of which they have been accused". It called for an amnesty for all six academics, or for each to receive a new trial "that meets international fair trial standards".

"It's terrible — though I hadn't been optimistic," says Carol Corillon, executive director of the International Human Rights Network of Academies and Scholarly Societies in Washington DC and co-author of the report along with Zenner and Peter Diamond, a Nobel prizewinning economist at the Massachusetts Institute of Technology in Cambridge. "It is a miscarriage of justice — there was no evidence at all against any of them. We have not had time to think of our next step, but we never, ever drop a case."

"This international delegation reflects what many of us believe — that there were many irregularities," says economic historian Şevket Pamuk, who is foreign secretary of the Bilim Academy in Istanbul, Turkey's independent national academy of sciences. "Many have argued that evidence was fabricated."

The scientists were sentenced as part of a trial code-named Ergenekon, in which 275 people, mostly military personnel, were accused of participating in a purported 'deep-state' network that the government believed had intended to facilitate a military coup. Observers describe the trial as a stand-off between a secularist old guard, which held power until 2003, and the current mildly Islamic government of Recep Tayyp Erdoğan. Erdoğan has increased his majority in parliament since first taking power, and is becoming more confident, says Pamuk. Critics suspect that the prominent academics in ▶

▶ the Ergenekon trial are being punished for stances unrelated to terrorism: all six are staunch secularists and have defended secularism by, for example, seeking to uphold a ban on headscarves in Turkey's universities.

The delegation's report also describes the plight of scientists, including one of the six just sentenced, accused in three further major political trials — known as Sledgehammer, the KCK Operations and the Postmodern Coup. All the political trials had been assigned to special anti-terrorism courts, but these were abolished last year following criticisms that they ignored evidence. Many observers believe that the trials, which the report described as "highly irregular", were used as an excuse to round up and silence government critics.

The Sledgehammer trial, which involved 365 people charged with attempting a military coup in 2003, ended last year. Industrial engineer Faruk Yarman, one of just two civilians to be charged, was sentenced to 13 years in prison. The report calls for his release.

Political scientist Büşra Ersanlı was arrested in 2011 as part of the KCK Operations, and charged with membership of a violent Kurdish-rights organization. She was released from pre-trial detention in July 2012; the report calls for a fair and expeditious trial for her.

And Kemal Gürüz, one of the six scientists sentenced this week, was arrested and detained in June last year as part of the Postmodern Coup trial (see *Nature* http://doi.org/h47; 2012).

A chemical engineer who was head of Turkey's Council of Higher Education from 1995 to 2003, Gürüz had been a vociferous proponent of the headscarf ban. He attempted suicide in prison in June this year. This week, he was sentenced to 13 years and 11 months in prison in the Ergenekon trial; he is still awaiting trial under the Postmodern Coup.

Pamuk, who holds joint positions at the Bosphorus University in Istanbul and the London School of Economics, says that many academics believe scientists such as Gürüz have been drawn into terrorism trials for reasons of revenge. "Many were university rectors," he says. "When they were powerful, they may have offended those who are now close to this government and are now in a position to retaliate." Pamuk expects all the scientists sentenced in the Ergenekon trial to appeal, but says that they are unlikely to win.

Guniz Gürüz, a chemical engineer at the Middle East Technical University in Ankara and Kemal Gürüz's wife, told *Nature* that pressures on the family have been extreme. "We are going to appeal," she said through tears after hearing his long sentence, adding, "Kemal has never had a proper explanation for why he was detained." ∎

Much of the discussion about the humane killing of research animals centres on rodents.

PASCAL GOETGHELUCK/SPL

ANIMAL RESEARCH

Best way to kill lab animals sought

Researchers debate most humane methods of dispatch.

BY DANIEL CRESSEY

Killing research animals is one of the most unpleasant tasks in science, and it is imperative to do it as humanely as possible. But researchers who study animal welfare and euthanasia are growing increasingly concerned that widely used techniques are not the least painful and least stressful available. This week, experts from across the world will gather in Newcastle upon Tyne, UK, to debate the evidence and try to reach a consensus.

"There are lots of assumptions made about the humaneness of various techniques for euthanizing animals," says Penny Hawkins, deputy head of the research animals department at the Royal Society for the Prevention of Cruelty to Animals, a charity based in Southwater, UK. "Sometimes an animal might not appear to be suffering, but might be conscious and suffering."

Much of the debate centres on rodents, which make up the vast majority of research animals. Current techniques for killing them include inhalation methods — such as chambers that fill with carbon dioxide or anaesthetic gases — and injecting barbiturates. Physical methods include cervical dislocation (breaking of the neck), or decapitation with specialist rodent guillotines (see 'Methods used to kill lab rats').

Experts hotly debate which method is

PROS AND CONS
Methods used to kill lab rats

Some methods recommended by the American Veterinary Medical Association.

Barbiturate injection: Fast-acting, but injection may cause pain.

Inhaled anaesthetic (halothane, isoflurane, sevoflurane or desflurane): Useful when restraint of animal is difficult.

Carbon dioxide: Acceptable, but chamber must be filled over several minutes and not pre-filled. Death to be verified afterwards or ensured by physical method.

Cervical dislocation: Causes rapid death, but skill must be learned.

Decapitation: Useful when tissues must be free of euthanasia chemicals.

Unacceptable: Nitrous oxide alone; nitrogen or argon asphyxiation (unless animals already anaesthetized); opioids.

preferable. The most-discussed question at the meeting is likely to be about the use of CO_2.

"People do still worry about CO_2 and it's still almost certainly the most widely used method [for killing rodents]," says Huw Golledge, who studies the anaesthesia of lab animals at Newcastle University, UK. Golledge organized the meeting, which is backed by the London-based National Centre for the Replacement, Refinement and Reduction of Animals in Research. Its aim is to update a 2006 consensus document produced by international experts to give guidance to researchers working with animals.

CO_2 is used to make rodents unconscious. They are then killed by either asphyxiating them with the gas, or by another method. But increasingly, studies suggest rodents find CO_2 stressful.

Evidence for this comes mainly from 'aversion studies'. A key study by animal welfare researcher Daniel Weary's group at the University of British Columbia in Vancouver, Canada, shows that albino rats will move away from a dark compartment filling with CO_2 into a brightly lit box, despite disliking bright lights. The study found that they were less likely to move away from isoflurane, also used in euthanasia (D. Wong *et al. Biol. Lett.* http://doi.org/ncv; 2012).

Other evidence is contradictory (H. Valentine *et al. J. Am. Assoc. Lab. Anim. Sci.* **51,** 50–57; 2012), but Weary is firm in his beliefs. "Our own results indicate CO_2 is highly aversive," he says.

There are also question marks over physical methods. Performed perfectly with animals accustomed to being handled, cervical dislocation may be the best method, but it may not be practical for killing large numbers of rodents.

The issues are even more uncertain for the new animal models that scientists are pursuing. For example, a huge increase in the use of zebrafish has put them on the meeting's agenda. Although much progress has been made with lab rodents, says Weary, "there's been much less work on fish welfare in general".

Widely used guidelines on animal euthanasia from the American Veterinary Medical Association (AVMA) in Schaumburg, Illinois, were updated earlier this year, in part to adapt to changes in the animals used in labs, with zebrafish guidance one of the additions. The association says it expects the lab animals section of these guidelines to continue to expand.

Some of this guidance comes with regulatory teeth. Later this year, the US National Institutes of Health, which funds biomedical research, says that it expects "full implementation" of the AVMA guidelines, with previously approved projects reviewed using them.

Regulation is also driving more unusual animals onto the agenda. New legislation on the treatment of laboratory animals is currently being incorporated into the laws of European Union member states and will cover cephalopods, in some nations for the first time (see *Nature* http://doi.org/fk65pb; 2011). ■

Squeezed light mutes quantum noise

Silicon zip reduces energy fluctuations in laser beams to improve sensitivity of optical motion sensors.

BY DEVIN POWELL

Oskar Painter spends his time carving silicon blocks into shapes that interact with light in strange ways. The latest of these — what Painter calls "Legos for adults" — squeezes the light from laser beams to push the limits of what they can measure.

By eliminating some of the noise caused by quantum effects, researchers can use squeezed light to illuminate movements too small to see with normal light. Painter's silicon sculpture, built on a microchip, could boost the sensitivity of sensors that use lasers to monitor motion, such as the gyroscopes that keep track of an aircraft's orientation.

"We have an opportunity to push the performance of these sensors by orders of magnitude," says Painter, an applied physicist at the California Institute of Technology in Pasadena, whose team reports how it squeezes light on page 185 (ref. 1).

All light is plagued by quantum noise, especially at the low powers typically required by sensors. These energy fluctuations blur the defined peaks of classical light waves, fundamentally limiting the precision of measurements.

Squeezing the light can suppress some noise, but Heisenberg's uncertainty principle demands a trade-off. A squeeze that reduces noise in one dimension — the height of a light wave's peaks, for instance — must be balanced by a stretch that adds noise in another, such as the distance between the peaks. Researchers therefore have to match the direction of the squeezing to the direction of the measurement.

Efforts to put light-squeezing to use have so far focused on gravitational-wave detectors, which search for faint ripples in spacetime by timing laser beams as they bounce between mirrors 4 kilometres apart. Passing ripples should stretch or compress the laser beams ever so slightly. But measurements with normal laser light are limited by quantum noise, and have so far failed to detect any disturbances attributable to gravitational waves.

Hoping to improve the next generation of measurements, researchers at the Laser Interferometer Gravitational-Wave Observatory (LIGO) in Hanford, Washington, added a dose of squeezed light by passing laser light through a crystal. In July, they reported that they had achieved a sensitivity better than the standard limit imposed by quantum noise[2]. This represents a step towards the ultimate goal of doubling LIGO's sensitivity, says team member Nergis Mavalvala, a physicist at the Massachusetts Institute of Technology in Cambridge. "We have to work hard to strip the noise out of the light," she says.

Painter's silicon device potentially offers a simpler way to squeeze light, although only at frequencies too high to be useful for gravitational-wave detectors. The device looks like a zip; photons bouncing around between its two arms push them apart with a force dictated by the amount of noise in the light. As the size of the gap changes, the zip tunes the frequency of the light — just as a finger sliding along a guitar string changes the pitch of the sound produced — and squeezes out some of the fluctuations.

> "We have to work hard to strip the noise out of the light."

The prototype tends to leak light, so it can suppress only about 5% of the noise. "The absolute level of squeezing is relatively low," says Warwick Bowen, a physicist at the University of Queensland in Brisbane, Australia. Painter says he will next be working with higher-quality zips, which could cut out as much as 90% of the noise.

But he will have some competition. A team at JILA in Boulder, Colorado, a joint institute of the University of Colorado and the US National Institute of Standards and Technology, has already created a vibrating silicon nitride membrane that boasts a 32% reduction in noise. JILA physicist Cindy Regal and her colleagues will report their work in a paper under review at *Physical Review X* (ref. 3). "It has been technically challenging to get to this regime," says Regal. ■

1. Safavi-Naeini, A. H. *et al. Nature* **500,** 185–189 (2013).
2. The LIGO Scientific Collaboration *Nature Photon.* **7,** 613–619 (2013).
3. Purdy, T. P., Yu, P.-L., Peterson, R. W., Kampel, N. S. & Regal, C. A. Preprint at http://arxiv.org/abs/1306.1268 (2013).

Henrietta Lacks with her husband David.

ETHICS

Deal done over HeLa cell line

Family of Henrietta Lacks agrees to release of genomic data.

BY EWEN CALLAWAY

Deborah Lacks wanted answers. In 1974, she asked a leading medical geneticist to tell her about HeLa cells, a tissue-culture cell line derived from the cancer that had killed her mother Henrietta in 1951. The researcher, who was collecting blood from the Lacks family to map HeLa genes, autographed a medical textbook he had written and said that everything she needed to know lay within its dense pages.

It would be more than 30 years before the family got a better explanation.

Now the director of the US National Institutes of Health (NIH), Francis Collins, is trying to make up for decades of slights. Over the past four months, he has met Lacks family members to answer questions and to discuss what should be done with genome data from their matriarch's cell line.

"We wanted to get a better understanding of what information was going to be out there about Henrietta, and what information was going to be out there about us," says Henrietta's grandson David Lacks Jr. (Deborah Lacks died in 2009.) On 7 August, Collins announced that the family has endorsed case-by-case release of the information, subject to approval by a

committee that will include family members (see page 141).

The consensual approach is a sea change from the dismissive treatment of the past, says Rebecca Skloot, the journalist who recounted the scene between Deborah Lacks and the researcher in her 2010 book *The Immortal Life of Henrietta Lacks*. "It was the first time in the very long history of HeLa cells that any scientists have sat down and devoted complete attention to explaining to the family what was going on," she says (see 'The Lacks legacy').

The agreement allows the publication of a US government-funded HeLa genome sequence as well as the re-release of data that were pulled from public view soon after publication in March because of the family's concerns. *Nature*'s News team learned of the negotiations last month but agreed to delay coverage so as not to impede the talks. Brokered during meetings at Johns Hopkins School of Medicine in Baltimore, Maryland, the deal rekindles debates over consent and ownership of tissues, and data that arise from their study, at a time when the NIH is updating such rules.

The HeLa cell line was established in 1951 from a biopsy of a cervical tumour taken from Henrietta Lacks, a working-class African-American woman living near Baltimore. The cells were taken without the knowledge or permission of her or her family, and they became the first human cells to grow well in a lab. They contributed to the development of a polio vaccine, the discovery of human telomerase and countless other advances. A PubMed search for 'HeLa' turns up more than 75,000 papers. "My lab is growing HeLa cells today," Collins told *Nature* in an interview on the NIH campus in Bethesda, Maryland. "We're using them for all kinds of gene-expression experiments, as is almost every molecular-biology lab."

On 11 March, weeks before Collins drove to Baltimore to meet the Lacks family for the first time, a team led by Lars Steinmetz at the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany, published a paper called 'The genomic and transcriptomic landscape of a HeLa cell line' (J. J. M. Landry *et al. Genes Genomes Genet.* http://dx.doi.org/10.1534g3.113.005777; 2013). News coverage (see go.nature.com/inxzuw) noted the link to Henrietta Lacks, but not privacy concerns.

Skloot, in a later article for *The New York Times*, made clear that family members were unhappy that — yet again — they had not been consulted. "I think it's private information," Henrietta's granddaughter Jeri Lacks-Whye told *Nature*. "I look at it as though these are my grandmother's medical records that are just out there for the world to see." The EMBL team removed the data from public access, and hoped that a solution could be reached.

As the controversy erupted, *Nature* was preparing to publish an even more detailed sequence of the HeLa genome, according to senior author Jay Shendure, a genome scientist

at the University of Washington in Seattle. His team, funded by the NIH, started decoding HeLa DNA in 2011, as part of an effort to develop new sequencing techniques. They also hoped that the genome would be useful for other researchers, a motivation shared by the EMBL team. They submitted their paper to *Nature* in November 2012.

The paper's reviewers did not raise privacy concerns before recommending it for publication; nor did *Nature*, Shendure says. He considered contacting the Lacks family before publication, and restricting access. "Figuring out how to reach out to the family was very much on the table when events overtook us."

After Skloot's article on the EMBL paper came out in March, Collins learned about Shendure's NIH-funded project. He saw an opportunity. He was already at work reforming the rules that govern research on human subjects. "It looked as if this was a moment to get everybody in the same room," he says.

And so, on the evening of 8 April, Collins met a group of Henrietta Lacks' children and grandchildren for dinner and discussion at the Johns Hopkins campus. Along with Collins was his chief adviser and two mediators from the university. Skloot phoned in to the meeting, which was to be the first of three.

Collins says that family members told him how unsettling it had been to learn about HeLa cells decades after Lacks died. They peppered Collins with questions about genetic sequencing and how Lacks' cells had been used. "I felt like I was taking 'Biology 101'," says Lacks-Whye. Collins told them that Shendure's team might have identified the genetic change that made their grandmother's tumour so aggressive and HeLa cells so prolific. The NIH later put the family in touch with experts in clinical genetics who told them what health information could be gleaned from the genome, and the NIH offered to help family members have their own genomes sequenced and interpreted.

Collins says that he did not pressure the family to agree to the release of the HeLa genome data; he was open to leaving the NIH-funded work unpublished. But he told the family that it would be impossible to keep the data locked away. NIH researchers had calculated that 400 genomes' worth of HeLa data are already publicly available in piecemeal form — parts of projects such as the Encyclopedia of DNA Elements — and that scientists in thousands of labs around the world could easily and cheaply sequence the cell line themselves.

Some Lacks family members raised the possibility of financial compensation, Collins says. Directly paying the family was not on the table, but he and his advisers tried to think of other ways the family could benefit, such as patenting a genetic test for cancer based on HeLa-cell mutations. They could not think of any. But they could at least reassure the family that others would not make a quick buck from their grandmother's genome, because

THE LACKS LEGACY

Story of the world's most widely used human biological research tissue.
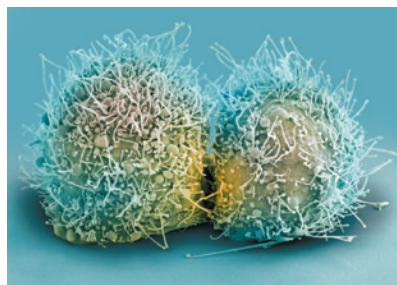
1951 Biopsy of Henrietta Lacks' tumour collected without her knowledge or consent. HeLa cell line soon established.

1971 The journal *Obstetrics and Gynecology* names Henrietta Lacks as HeLa source; word later spreads in *Nature, Science* and mainstream press.

1973 Lacks family members learn about HeLa cells (pictured). Scientists later collect their blood to map HeLa genes, without proper informed consent.

1996 Lacks family honoured at the first annual HeLa Cancer Control Symposium, organized by former student of scientist who isolated HeLa cells.

2013 HeLa genome published without knowledge of the family, which later endorses restricted access to HeLa genome data.

the US Supreme Court had this year ruled that unmodified genes could not be patented. Lacks-Whye says that the family does not want to dwell on money — and that her father has often said he "feels compensated by knowing what his mother has been doing for the world".

In the end, the family decided that it wanted the data to be available under a restricted-access system similar to the NIH dbGaP database, which links individuals' genetic make-up to traits and diseases. Researchers would apply for permission to acquire the data and agree to use them for biomedical research only, and

would not contact Lacks family members. A committee that includes family members will handle requests, and papers that use the data will recognize Henrietta Lacks and her kin. The first of these papers, the NIH-funded paper, is published in this issue (see page 207).

In discussing HeLa cells and the agreement forged with the family, Collins and others often use the word "unique". No other human sample matches the cell line for ubiquity, notoriety or celebrity (Oprah Winfrey is producing a film based on the story). The NIH does not see the deal with the family as a guide to handling other human samples. "It's not going to be a precedent," says Collins' chief adviser Kathy Hudson.

But it will probably inform other cases, she adds. The US government is redrafting rules that govern the relationship between federally funded researchers and participants. New rules aim to give subjects greater say in how their tissues and personal data are used. "Going forward, I'm very much of the mind that the most appropriate way to show respect for persons is to ask," Collins says. "Ask people, 'Are you comfortable having this specimen used for future genomic research for a broad range of biomedical applications?' — if they say no, no means no."

As for the myriad other tissues out there that were obtained without consent, Collins says that it would slow science too much to ban their use. Laura Rodriguez, a policy official at the NIH who works on guidelines for genome sequencing, says that there is a low risk of donors of such samples being identified. But in January, researchers working on a genomics project showed that it is possible to identify anonymous participants — and their families — by cross-referencing their genomes with genealogy DNA databases.

Hank Greely, a biotechnology lawyer at Stanford University in California who has advised the EMBL group, says the HeLa agreement is a "good solution", but applying it to other unconsented cell lines and data would be unwieldy and impractical. "The one thing we really should be doing is making sure everything we collect from here into the future is acceptable."

Lacks-Whye has similar advice. Researchers can make major breakthroughs, she says, while still respecting the wishes of patients and their families. "Have them involved," she says. "That's not only for HeLa sequences, but anybody who participates in research." ■ **SEE EDITORIAL P.121**

TOP STORY

Climatic swings have incited human conflict throughout history
go.nature.com/b5nrl6

MORE NEWS

● Synthetic molecule fights TB in mice go.nature.com/ilvazz
● Genes linked to human ability to smell four scents go.nature.com/seiwvy
● Undernourishment during development reduces fitness later in life go.nature.com/lxigvk

X-rays emitted by swirling disks of matter hint at the speed at which a supermassive black hole spins.

ASTROPHYSICS

Spin rate of black holes pinned down

Calculation offers way to probe galactic evolution.

BY EUGENIE SAMUEL REICH

Black holes can be described by just two fundamental characteristics: mass and spin. Astronomers have been able to measure the objects' mass for decades, by looking for gravitational effects on the orbits of nearby stars. But measuring spin, which records the angular momentum of the matter that falls into the holes, has proved troublesome, particularly for the supermassive black holes that lie at the centres of galaxies. No light emanates from the black holes' spinning event horizons, so astronomers instead look for proxies that emit X-rays, such as the swirling disks of matter that feed into some holes.

Such indirect spin measurements have now been made for 19 supermassive black holes for which the mass is also well known (see 'Spin off'). On 29 July, astronomers reported that they had calculated the spin of another supermassive black hole, using a new technique that, although unproven, provides an alternative way to target the elusive quantity. "There's a significant number of us who think we are getting a coherent picture of black hole spin," says Andrew Fabian, an astronomer at the University of Cambridge, UK.

The conventional method used to measure spin dates to 1995, although this has been controversial until recently. It relies on the detection of X-rays emitted from the corona, a spherical halo of hot, ionized gas that sits just above and below the plane of the black hole's accretion disk. Some of these X-rays bounce off the disk and travel towards Earth. In them, astronomers can sometimes discern a prominent emission line characteristic of iron. The higher the black hole's spin, the closer the accretion disk can get to the black hole's event horizon, and the more the strong gravity can distort the iron line, spreading it over a wider range of X-ray energies.

Scepticism about the method is beginning to lift. In February, astronomers published spin calculations (G. Risaliti *et al. Nature* **494,**

449–451; 2013) that used data from NASA's NuSTAR mission, which was launched last year (see *Nature* **483,** 255; 2012). Study leader Guido Risaliti, an astronomer at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, says that NuSTAR provides access to higher-energy X-rays, which has allowed researchers to clarify the influence of the black hole's gravity on the iron line. These rays are less susceptible than lower-energy X-rays to absorption by clouds of gas between the black hole and Earth, which some had speculated was the real cause of the distortion.

In the latest study, astronomers calculated spin more directly (C. Done *et al. Mon. Not. R. Astron. Soc.* http://doi.org/nc2; 2013). They found a black hole some 150 million parsecs away with a mass 10 million times that of the Sun. Using the European Space Agency's XMM-Newton satellite, they focused not on the iron line but on fainter, lower-energy X-rays emitted directly from the accretion disk. The spectral shape of these X-rays offers indirect information about the temperature of the innermost part of the disk — and the temperature of this material is, in turn, related to the distance from the event horizon and the speed at which the black hole is spinning. The calculations suggest that, at most, the black hole is spinning at 86% of the speed of light.

Study leader Chris Done, an astronomer at Durham University, UK, thinks her result casts doubt on spin measurements made using the iron line, because those results tend to come in at above 90%. "We're on the very edge of what we can do," she says. "We have different methods and we'd like them to agree." Others argue that the differences in results may reflect a genuine variation between supermassive black holes, and suggest that spin may vary with mass, or over cosmic time.

Much is at stake. If the spins of supermassive black holes are as high as some have found using the iron line, then these black holes are likely to have formed from rare, major mergers between colliding galaxies, in which a large quantity of material falls into the central black hole from one direction. If the spins are lower, as Done suggests, then the black holes may have formed from many minor mergers, with bite-sized lumps of material coming from various directions. The distribution of black hole spins could therefore inform researchers about the history of galactic evolution, particularly if astronomers can eventually chart the change in spin over cosmic time by looking at ever-more-distant black holes.

Astronomers also want to understand whether spins power the jets of material that spew out from some black holes. But they cannot address these questions while disagreement remains over the spin measurement techniques, says Risaliti. He is optimistic that further X-ray observations will resolve the controversy. "It's a long way to go, but this is the beginning," he says. ∎

SPIN OFF

Some supermassive black holes spin at more than 90% of the speed of light, which suggests that they gained their mass through major galactic mergers.
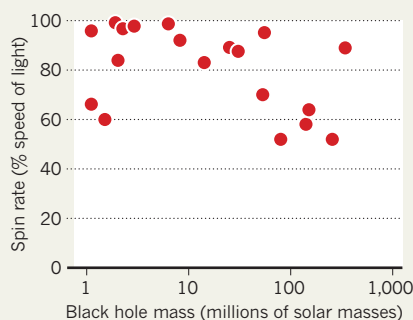
A LINE IN THE SANDS

The scientific community is sharply divided over the proposed Keystone XL pipeline from Canada's tar sands.

BY JEFF TOLLEFSON

Environmental activist Bill McKibben has called it the "fuse to the biggest carbon bomb on the planet". Famed US climate researcher James Hansen has warned that it would unleash a "monster". And protestors have chained themselves to the White House fence, declaring that it would feed a nasty fossil-fuel addiction and enrich the oil industry while dooming the global climate.

The object of all that ire, the Keystone XL pipeline, is designed to carry crude oil some 1,900 kilometres from the tar sands of Alberta, Canada, to the US Midwest, where it will link into a network of pipelines supplying refineries on the Gulf of Mexico. Proponents say that it would provide North America with a secure source of energy and reduce dependence on overseas oil. But for environmentalists frustrated by a stalemate in Congress and repeated failures to secure an aggressive international climate treaty, the pipeline has become a key battle — one that they hope will trigger a popular uprising against unbridled fossil-fuel development.

The issue has also divided the scientific community. Many climate and energy researchers have lined up with environmentalists to oppose what is by all accounts a dirty source of petroleum: emissions from extracting and burning tar-sands oil in the United States are 14–20% higher than the country's average oil emissions. But other researchers say that the Keystone controversy is diverting attention from issues that would have much greater impact on greenhouse-gas emissions, such as the use of coal.

Some experts find themselves on both sides. "I'm of two minds," says David Keith, a Canadian climate scientist who is now at Harvard University in Cambridge, Massachusetts. "The extreme statements — that this is 'game over' for the planet — are clearly not intellectually true, but I am completely against Keystone, both as an Albertan and somebody who cares about the climate."

SIGNIFICANT FIGURES

The pipeline's future rests with US President Barack Obama, who declared in June that Keystone would serve the national interest only if it "does not significantly exacerbate the problem of carbon pollution". The debate now centres on the definition of 'significantly', which requires a bit of context.

Canada has an estimated 170 billion barrels of dense, viscous oil locked up in deposits of loose sandstone in Alberta. These tar sands, or oil sands, produced 1.8 million barrels of oil per day in 2012, and that figure is projected to nearly triple by 2030. More than two-thirds of that oil makes its way via pipelines to the United States, where it accounts for around 7% of US oil consumption. But the pipelines are reaching capacity, so companies looking to increase production must first figure out how to get their product out of Canada. Keystone is just the first step and would eventually carry some 730,000 barrels per day from the tar sands to US refineries. To meet the industry's production forecast, however, at least three more pipelines of comparable capacity will be needed.

A draft environmental-impact statement by the US Department of

The tar sands of Alberta supply 7% of the oil used in the United States.

JIRI REZAC/GREENPEACE/EYEVINE

State found that halting the Keystone pipeline would have a minimal impact on the development of the tar sands, because oil companies would find other ways to get their product to market. In the short term, that means shipping by rail, which would increase emissions. But the state department's conclusions have come under fire from environmentalists as well as the US Environmental Protection Agency, which urged the state department to conduct "a more careful review" of the economic analysis for its final assessment.

There is cause for scepticism about the alternatives to the Keystone pipeline. Although companies are increasingly using rail to ship conventional oil out of North Dakota, where production is booming, that option is less attractive for the oil sands. Rail cars carrying tar-sands oil cost more to run because they must be heated, and they cannot carry as much, because the oil is much heavier than typical crude. Trains from Alberta must also travel much farther to reach the Gulf coast. The combination of factors increases prices by about US$20 per barrel.

BARRELLING FORWARD

Other proposed pipelines could transport the oil, but they also face challenges. Last week TransCanada, the company behind the Keystone XL plan, said that it wants to build a $12-billion pipeline to the country's Atlantic coast, but environmentalists concerned about oil spills have vowed to block that route. And British Columbia's ruling party took a stance in May against a planned pipeline to Canada's Pacific coast.

"Really what you are left with is that the Keystone pipeline is the only route forward" to increase production from the tar sands, says Susan Casey-Lefkowitz, director of the international programme for the Natural Resources Defense Council in Washington DC. And because increasing production necessarily increases emissions, she says, Keystone "fails the president's climate test".

But the matter is far from settled. The oil industry has already begun shipping conventional oil out of Alberta by rail, and IHS Cambridge Energy Research Associates, a consulting firm in Englewood, Colorado, projects that tar-sands producers will employ rail if no pipelines are built. "We think it's economic, and we think it will grow in the absence of pipelines," says Jackie Forrest, senior director for global oil at the firm. Others argue that the industry would push to build other pipelines if Keystone were to fail.

The environmental impact must also be weighed against issues such as safety and energy security, says Andrew Weaver, a climate scientist at the University of Victoria in British Columbia, who was elected this year to the province's parliament as a member of the Green Party. He refuses to weigh in on whether the pipeline should be built, saying that the decision rests with the United States. But he calls the argument for North American energy security "quite compelling". And he cites the crash last month of an oil-transport train in Quebec, as evidence that the potential for human error is higher with rail than with pipelines. "I think it's mad that we are burning all of our oil," Weaver says, "but we've got to put it into perspective."

In 2012, Weaver sought to do just that. He and a student calculated what would happen to global temperatures if the tar sands were fully developed. The proven reserves — those that could be developed with known technologies — make up roughly 11% of the global total for oil, and Weaver's model suggested that full development would boost the average global temperature by just 0.03 degrees Celsius (N. C. Swart and A. J. Weaver *Nature Clim. Change* **2,** 134–136; 2012). Weaver says that the initial focus should be on coal, which he found would have 30 times the climate impact of oil if the world burned all proven coal reserves.

"As a serious strategy for dealing with climate, blocking Keystone is a waste of time," says David Victor, a climate-policy expert at the
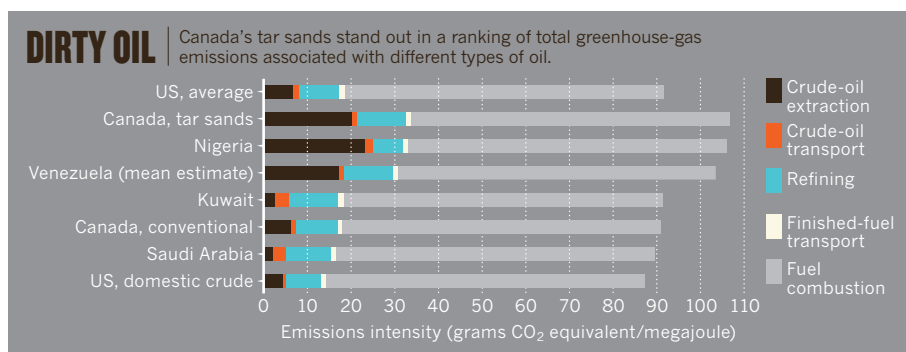
University of California, San Diego. "But as a strategy for arousing passion, it is dynamite."

Well aware that their future prosperity may depend on it, producers as well as the government of Alberta — which hauled in about $4.3 billion in royalties from the oil sands in the 2011–12 fiscal year — say that they are cleaning up operations there. Within Canada much of the concern has focused on local pollution from the mining operations, which spew exhaust and toxic chemicals into the atmosphere and leave behind large wastewater ponds. But companies are also working to trim the overall greenhouse-gas emissions associated with production.

Environment Canada, the country's environment agency, claims that tar-sand producers reduced their emissions by 26% per barrel of oil between 1990 and 2010. But emissions are poised to increase in the coming years as companies probe deeper into the earth. The industry is up against geology. Having depleted many of the tar-sand deposits accessible through surface mining, companies are exploiting deeper formations by injecting steam into the rock layers to liquefy and produce the oil. Producing steam requires natural gas, which can increase emissions by up to 30% compared with the surface-mining process (see 'Dirty oil'). Those deeper deposits now account for roughly half of oil-sands production, but they make up 80% of proven reserves, so their share of production will only climb higher.

Alberta is investing more than $700 million in a $1.35-billion demonstration project that would capture and bury up to 1.2 million tonnes of carbon dioxide annually from a facility that upgrades bitumen, the tar-like product from the oil sands, into crude oil for shipment. On its own, however, that project would not have a significant impact on emissions. More generally, Alberta enacted a law in 2007 requiring all major emitters in the province to reduce their annual emissions intensity — a measure of emissions relative to production — by 12% or face a levy of $14 for every tonne of emissions in excess of that target. That levy has raised $376 million for clean-energy investments to date, nearly $79 million of which has been invested in projects related to the tar sands.

Yet there is no way to cleanly produce oil from the tar sands — or

DIRTY OIL | Canada's tar sands stand out in a ranking of total greenhouse-gas emissions associated with different types of oil.

US, average
Canada, tar sands
Nigeria
Venezuela (mean estimate)
Kuwait
Canada, conventional
Saudi Arabia
US, domestic crude

Emissions intensity (grams CO$_2$ equivalent/megajoule)
0 10 20 30 40 50 60 70 80 90 100 110

- Crude-oil extraction
- Crude-oil transport
- Refining
- Finished-fuel transport
- Fuel combustion

from anywhere else. "The way that you drive down emissions in the transportation sector is by driving less, by becoming more efficient, and then by changing your fuels," says Michael Levi, an energy-policy fellow at the Council on Foreign Relations in New York.

Many researchers who have sided with environmentalists on Keystone acknowledge that the decision is mostly symbolic. But in the absence of other action, says Harvard's Keith, it is important to get people involved and to send industry a message that the world is moving towards cleaner fuels, not dirtier ones.

For Ken Caldeira, a climate researcher at the Carnegie Institution for Science in Stanford, California, it is a simple question of values. "I don't believe that whether the pipeline is built or not will have any detectable climate effect," he says. "The Obama administration needs to signal whether we are going to move toward zero-emission energy systems or whether we are going to move forward with last century's energy systems." ■

Jeff Tollefson *covers energy and environment for* Nature *in New York.*

METAMATERIAL WORLD

BY LEE BILLINGS
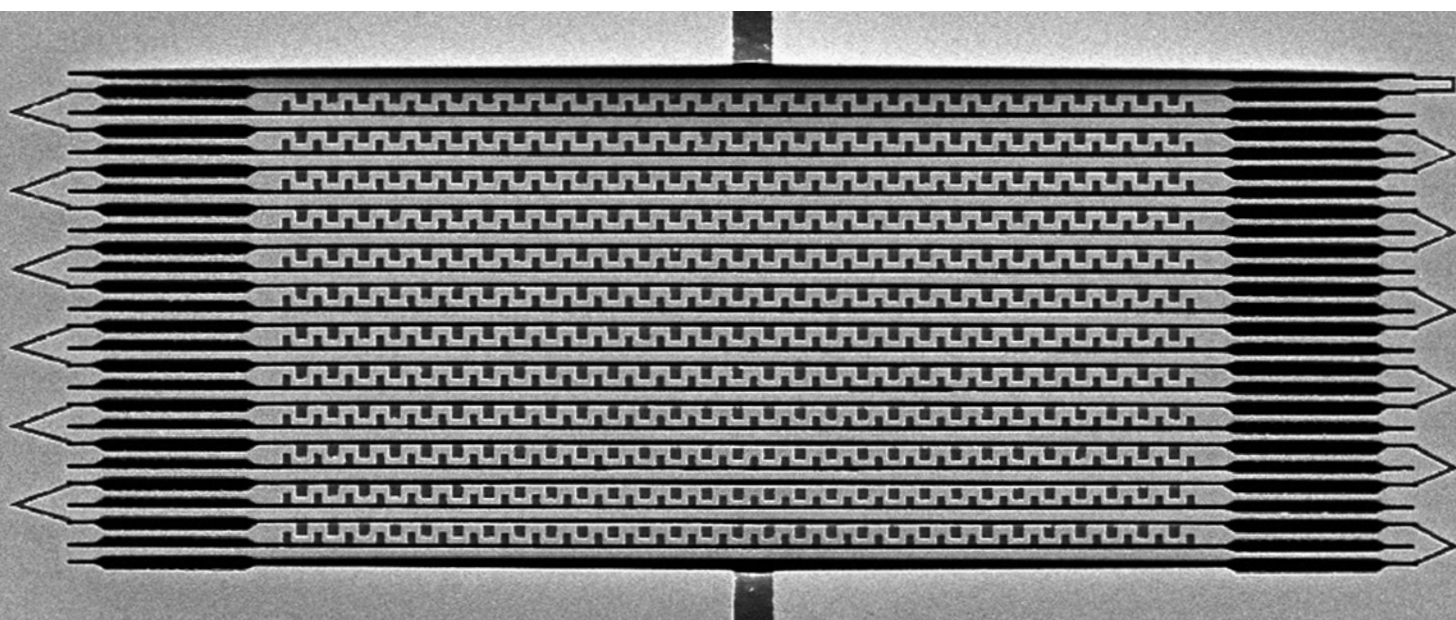
Engineered structures with bizarre optical properties are set to migrate out of the laboratory and into the marketplace.

Tom Driscoll would be happy if he never heard the phrase 'Harry Potter-style invisibility cloak' again. But he knows he will. The media can't seem to resist using it when they report the latest advances in metamaterials — arrays of minuscule 'elements' that bend, scatter, transmit or otherwise shape electromagnetic radiation in ways that no natural material can. It is true that metamaterials could, in principle, route light around objects and render them invisible, not unlike the cloak of a certain fictional wizard. And many metamaterials researchers are trying to make cloaking a reality, not least because the military has eagerly funded the development of such capabilities.

However, if such applications ever come to pass it will be decades from now. Technologies closer to

commercialization are of more interest to Driscoll, a physicist who oversees metamaterials commercialization at Intellectual Ventures, a patent-aggregation firm in Bellevue, Washington. Applications such as cheaper satellite communications, thinner smartphones and ultrafast optical data processing are "where metamaterials are poised to make a huge impact", he says.

Researchers still face some daunting challenges, he adds — notably, finding cheap ways to fabricate and manipulate metamaterial elements on a scale of nanometres. But the first metamaterial-based products are expected to come onto the market in a year or so. And, not long after that, Driscoll expects that average consumers will start to enjoy the benefits, such as faster, cheaper Internet connectivity on board planes and from mobile phones. Such applications, he says, will move from being the stuff of

peoples' fantasies "to becoming things they can't contemplate living without".

The first laboratory demonstration of a metamaterial was announced in 2000 by physicist David Smith and his colleagues at the University of California, San Diego[1]. Following up on theoretical work done in the 1990s by John Pendry of Imperial College London, these researchers showed that an array of tiny copper wires and rings had a negative refractive index for microwaves — meaning that microwave radiation flowing into the material is deflected in a direction opposite to that normally observed (see 'Wave engineering'). That triggered intense interest in metamaterials, in part because the ability to bend radiation in such a way had potential for creating invisibility cloaks.

Since then, Smith and others have explored a host of variations on the metamaterial idea, often looking to manipulate radiation in ways that have nothing to do with a negative refractive index. They have also moved beyond static arrays, devising techniques to change the way the elements are arranged, how they are shaped and how they respond to radiation. The resulting materials can do things such as turn from opaque to transparent or from red to blue — all at the flick of a switch.

MARKET MOVERS

In January, Smith, now at Duke University in Durham, North Carolina, took on a concurrent role as director of metamaterials commercialization efforts at Intellectual Ventures. "I felt that the time was right, and we didn't need to do any more science for some of these things," he says.

A test case may come as early as next year. Kymeta of Redmond, Washington, a spin-off from Intellectual Ventures, hopes to market a compact antenna that would be one of the first consumer-oriented products based on metamaterials. The relatively inexpensive device would carry broadband satellite communications to and from planes, trains, ships, cars and any other platform required to function in remote locations far from mobile networks.

At the heart of the antenna — the details of which are confidential — is a flat circuit board containing thousands of electronic metamaterial elements, each of which can have its properties changed in an instant by the device's internal software. This allows the antenna to track a satellite across the sky without having to maintain a specific orientation towards it, the way a standard dish antenna does. Instead, the antenna remains still while the software constantly adjusts the electrical properties of each individual metamaterial element. When this is done correctly, waves emitted from the elements will reinforce one another and propagate skywards only in the direction of the satellite; waves emitted in any other direction will cancel one another out and go nowhere. At the same time — and for much the same reason — the array will most readily pick up

signals if they are coming from the satellite.

This technology is more compact than alternatives such as dish antennas, says Smith. It offers "significant savings in terms of cost, weight and power draw". Kymeta has already performed demonstrations of this technology for investors and potential development partners. But Smith cautions that the company has yet to set a price for the antenna and that it must still work to bring production costs down while maintaining the strict performance standards that regulatory agencies demand for any device communicating with satellites.

Kymeta has shared so few details of its antenna that researchers say it is hard to offer an evaluation. But Smith is highly regarded in the field. If Kymeta brings the product to market, it may first offer its antenna for use on private jets and passenger planes. If buyers respond well, the company hopes to incorporate the technology into other product lines, such as portable, energy-efficient satellite-communication units for rescue workers or researchers in the field.
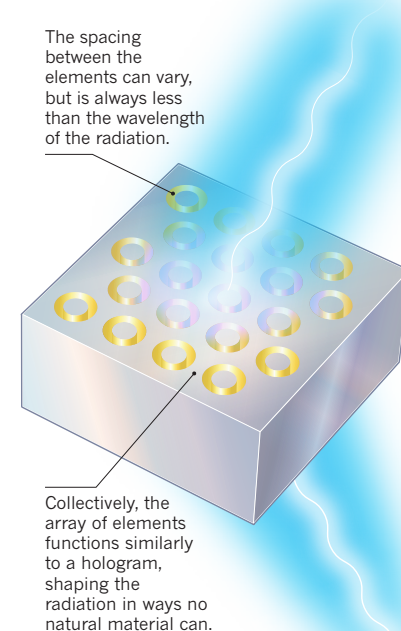
In January, Smith's group turned heads when it announced its demonstration of another metamaterial device: a camera that can create compressed microwave images without a lens or any moving parts[2]. One important application of the device might be to reduce the cost and complexity of airport security scanners.

In their current form, these scanners have to physically sweep a microwave sensor over and around the subject. This produces an unwieldy amount of data that has to be stored before it is processed into an image. The Duke group's device requires very little data storage. It takes numerous snapshots by sending beams of microwaves of multiple wavelengths across the target at about ten times per second. When the microwaves are reflected back by the subject, they fall on a thin strip of square copper metamaterial elements, each of which can be tuned to block or let through reflected radiation. The resulting pattern of opaque and transparent elements can be varied very rapidly, with each configuration transmitting a simplified snapshot of a scanned object into a single sensor. The sensor measures the total intensity of radiation from each snapshot, then outputs a stream of numbers that can be digitally processed to reconstruct a highly compressed image of the subject.

This is admittedly just a first step: demonstrations carried out so far have been crude affairs restricted to two-dimensional images of simple metallic objects. Expanding it to three-dimensional images of complex objects remains a challenge. But if that challenge can be overcome, says Driscoll, airports could retire the bulky, expensive, slow booths that currently constitute security checkpoints, and instead use a larger number of thin, inexpensive metamaterial cameras hooked up to computers. Such a shift, Driscoll says, could extend security scanning to rooms, hallways, and corridors throughout airports and other sensitive facilities.

Wave engineering

Metamaterial elements scatter incoming radiation in very precise ways. They can be any shape; common examples include spheres, rings, crosses and chevrons. Their electromagnetic properties can often be changed by software.


The spacing between the elements can vary, but is always less than the wavelength of the radiation.

Collectively, the array of elements functions similarly to a hologram, shaping the radiation in ways no natural material can.

Example:
Negative index of refraction
Metamaterials can be engineered to bend radiation in a direction opposite to that observed in ordinary materials.

Positive refraction — Negative refraction
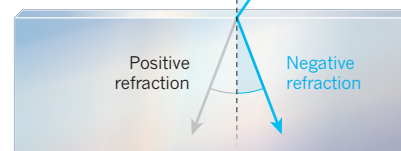
Application:
Invisibility cloak
A cloak made of a negative-index metamaterial can bend radiation around an object inside it, making that object seem invisible.

Radiation source

Metamaterial cloak
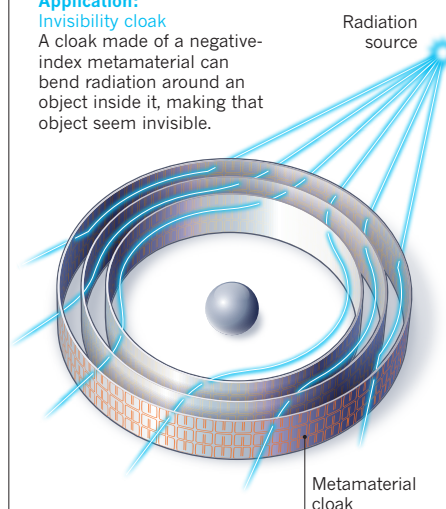
In the meantime, a key research goal for Smith and his group is the development of robust and marketable metamaterial devices that are not restricted to radio, microwave or infrared wavelengths. If the technologies could be made to work with visible light, they would become much more useful for applications such as fibre-optic communications or consumer-oriented cameras and displays.

"It won't be easy," cautions Stephane Larouche, a member of Smith's research team at Duke. For any given type of radiation, he explains, metamaterials can wield their exotic powers only if the elements are smaller and more closely spaced than the wavelength of that radiation. "So the shorter and shorter the wavelength we wish to use, the smaller each metamaterial element must be," says Larouche.

In the microwave and radio regions of the spectrum, this is relatively easy: wavelengths are measured in centimetres to metres. But an optical metamaterial's elements would have to measure considerably less than a micrometre. That is not impossible: today's high-performance microchips contain features only a few tens of nanometres across. But unlike those essentially static features, says Larouche, the metamaterial elements in many applications would need to incorporate ways for software to change their properties dynamically as needed. "Too often we have gorgeous ideas," he says, "but we have no way of fabricating them."

FLAT FOCUS

Despite these difficulties, workable designs for optical metamaterials have begun to emerge. One was published in March[3] by a group working under Nikolay Zheludev, a physicist at the University of Southampton, UK, who directs a research centre focused on metamaterials at Nanyang Technological University in Singapore. The team's device can greatly alter its ability to transmit or reflect optical wavelengths by means of nanometre-scale, electrically controlled metamaterial elements etched from gold film; it could one day serve as a switch in high-speed fibre-optic communications networks.

Meanwhile, because it is so hard to make and control three-dimensional metamaterial arrays at optical scales, some researchers are focusing on two-dimensional 'metasurfaces'. In August 2012, a group led by Federico Capasso at Harvard University in Cambridge, Massachusetts, unveiled a flat metamaterial lens that can focus infrared light to a point in much the same way as a glass lens[4]. "I don't want to claim absolute novelty in this," Capasso says, "but I believe we are the first group to so clearly put flat optics on the agenda for commercial applications."

A conventional lens relies on refraction to bend light to a point by passing it through varying thicknesses of glass. Capasso's lens passes light through a two-dimensional array of gold metamaterial elements carved out of a 60-nanometre-thick silicon wafer using electron-beam lithography techniques developed for the microchip industry. The elements are fixed, so cannot be tuned after fabrication. But by selecting a specific size and spacing during the manufacturing process, physicists can shape light of a chosen wavelength in exactly the right way to make it come to a point.

Capasso warns that commercial applications of such flat lenses are probably still a decade away. This is partly because silicon is a rigid and fragile substrate for etching the elements; researchers are looking at more robust and flexible alternatives that would be easier to handle on the production line. They are also looking for better ways to control the carving of the nanoscale elements, which has to be done very precisely.

But once the technology is mastered, says Capasso, one obvious application is in smartphone cameras. Lenses, along with batteries, are among the most stubborn limiting factors

> ## "Metamaterials are poised to make a huge impact."

in smartphone thickness, he says, speculating that a smartphone incorporating a flat camera lens could potentially be made "as thin as a credit card". The flat lens also avoids aberrations that plague glass lenses, such as the coloured 'fringes' created by the inability to focus all wavelengths to the same point. This means that Capasso's flat lens could also be used to make better, aberration-free microscopes.

As good as they might ultimately be, the flat lenses would still be subject to the diffraction limit, which dictates that no conventional lens can resolve details much smaller than the wavelength of the light that illuminates its target. This limit averages about 200 nanometres for visible light. But metamaterials offer a means of fabricating 'superlenses' that could surpass such limits, allowing researchers to see sub-wavelength details of target objects such as viruses or the ever-changing structures in living cells.

The key is to recognize that the missing details are still there, carried in 'evanescent' waves of reflected light that die away very rapidly with distance from the illuminated object. Normally, these waves have effectively vanished before they can be captured and focused by a lens. But a metamaterial superlens designed to be placed within tens of nanometres of an object can pick up and magnify these waves.

An early proof-of-concept superlens was demonstrated in 2005 by a group working under Xiang Zhang, a physicist at the University of California, Berkeley[5]. Zhang's group produced a simple metamaterial consisting of a 35-nanometre-thick layer of silver in a sandwich with nanoscale layers of chromium and plastic.

The team has since been working to refine the superlens concept; in 2007 the researchers advanced the idea by developing 'hyperlenses' from curved, nested layers of compounds such as silver, aluminium and quartz[6]. The lenses not only capture evanescent waves, but can also feed them into a conventional optical system. Ultimately, this could allow sub-wavelength details to be viewed through the eyepiece of a standard microscope. But the complex structure and behaviour of hyperlenses makes them difficult to manufacture and use in this way.

REVERSIBLE FOCUS

By pairing conventional optics with superlenses and hyperlenses based on metamaterials, Zhang hopes eventually to find applications far beyond microscopy. Just as these constructs can magnify sub-wavelength detail, they can also be run in reverse, directing beams of light into sub-wavelength focal points — a property of potentially revolutionary importance for fabricating minuscule structures using photolithography. If superlenses and hyperlenses can be harnessed for this purpose, the ultra-fine beams of light could be used to etch much smaller features than is possible today. This could greatly increase the density of data storage on optical drives, as well as the number of components that can be crammed onto computer chips.

Smith is cautious on that score, pointing out that hyperlenses and superlenses tend to dissipate substantially more of the light energy passing through them than other advanced lithographic techniques now in development. This, he says, makes them prime examples of "strong and compelling science that is not yet practical for any sort of product path" at optical wavelengths. But, he adds, Zhang's efforts are "heroic experiments that illustrate the potential of metamaterials in a fundamental way".

Zhang concedes that hyperlenses and superlenses are not yet ready for prime time, but believes there is plenty of room for ongoing research to change that situation in the coming years. "The economic impact could be huge," he says. "I am cautiously optimistic that metamaterials, superlenses and lithography will prove truly revolutionary. If people aren't too short-sighted, what we can do with metamaterials will be limited only by our imaginations." ∎

Lee Billings *is a freelance writer based in New York.*

1. Smith, D. R., Padilla, W. J., Vier, D. C., Nemat-Nasser, S. C. & Schultz, S. *Phys. Rev. Lett.* **84,** 4184–4187 (2000).
2. Hunt, J. *et al. Science* **339,** 310–313 (2013).
3. Ou, J.-Y., Plum, E. Zhang, J. & Zheludev, N. l. *Nature Nanotechnol.* **8,** 252–255 (2013).
4. Aieta, F. *et al. Nano Lett.* **12,** 4932–4936 (2012).
5. Fang, N., Lee, H., Sun, C. & Zhang, X. *Science* **308,** 534–537 (2005).
6. Liu, Z., Lee, H., Xiong, Y., Sun, C. & Zhang, X. *Science* **315,** 1686 (2007).

COMMENT

Henrietta Lacks' family gather around a historical marker dedicated to her in Virginia in 2011.

Family matters

Kathy L. Hudson and **Francis S. Collins** discuss how and why the US National Institutes of Health worked with the family of Henrietta Lacks, the unwitting source of the HeLa cell line, to craft an agreement for access to HeLa genome data.

In March, two of the most deeply held values in the medical-research community — public data-sharing and respect for research participants — collided when the genome of the ubiquitous cell line HeLa was published[1] and posted in a public database. Controversy ensued. The full sequence data could potentially uncover unwanted information about people whose identity is widely known: the family of the woman from whom this immortal line was derived 62 years ago, Henrietta Lacks.

So, since March, the US National Institutes of Health (NIH) in Bethesda, Maryland, has worked closely with Lacks' family. Together, we have crafted a path that addresses the family's concerns, including consent and privacy, while making the HeLa genomic sequence data available to scientists to further the family's commitment to biomedical research.

The agreement that we reached goes into effect this week. We hope that it, and its genesis, will spur broader discussions regarding consent for future use of biospecimens, with a goal of fostering true partnerships between researchers and research participants.

MEDICAL HISTORY

In 1951, physicians at Johns Hopkins Hospital in Baltimore, Maryland, took a biopsy from Henrietta Lacks, a 31-year-old African American woman who had an aggressive form of cervical cancer. This biospecimen was taken without her permission or knowledge; US regulations requiring consent ▶

LACKS FAMILY/VIRGINIA DEPT HISTORIC RESOURCES

were still decades away. The tissue sample gave rise to the first human cancer-cell line that could grow endlessly in culture, called HeLa. Henrietta died later that year, but her cells live on. Today, more than 60 years later, scientists around the world use HeLa cells for research on almost every disease. The story of Lacks' unwitting contribution to science, and the proud and poignant legacy it left for her descendants, is told in Rebecca Skloot's best-selling book, *The Immortal Life of Henrietta Lacks* (Crown, 2010), which is now being made into a film by Oprah Winfrey's production company.

The German research team that in March this year posted the HeLa genome on open-access databases available through the European Bioinformatics Institute and the NIH's National Center for Biotechnology Information did not violate any laws or rules. The action did, however, upset the Lacks family, and it drew criticism from many quarters[2]. The genome of these cells is not identical to Lacks' original genome. The cells carry the genetic modifications that allowed them to form a tumour and grow prolifically; and their passage in cell culture for more than six decades has led to other structural anomalies. Nonetheless, the sequence can reveal certain heritable aspects of Lacks' germline DNA, and can thus be used to draw inferences, admittedly of uncertain significance, about her descendants.

Within days, the European researchers removed the sequence from the public databases, to allow time for consideration of alternative approaches. Meanwhile, an NIH-funded research paper by Andrew Adey and colleagues on the genome sequence of a second HeLa line was in press at *Nature* (published in this issue; see page 207)[3]. *Nature* mandates that authors of research papers make their data publicly available online. Something needed to be done — and in partnership with the Lacks family.

WEIGHING THE OPTIONS

Over the past four months, with help from Skloot and academic leaders at Johns Hopkins, we met members of the Lacks family in Baltimore on three occasions. At their request, some family members also met separately with an NIH genetic counsellor and medical-genetics expert to learn more about what the data might say about family members, and the implications of having it in the public domain.

We talked at length with the family about the three options available for the full HeLa sequence data: first, making the sequence freely available, allowing anyone access at any time and for any use; second, placing the data in a controlled-access database, which would require researchers to apply to the NIH to use the data in a specific study and to agree to terms of use defined by a panel including

members of the Lacks family; or third, withholding the sequence and not making it available at all for research — an option that the NIH would have had difficulty supporting or implementing, philosophically and legally.

After much discussion, family members unanimously favoured the controlled-access option. This will allow them to be aware of and have a crucial role in the science that uses the HeLa genome. The NIH will help to implement this, but respecting the family's preferences has required (and will continue to require) cooperation and patience by many — including scientists, publishers, funders and scientific societies. The authors and publishers of both genome papers[1,3] have agreed to submit their data for controlled access (in the same way as for many other non-HeLa genome sequences) through the NIH's database of genotypes and phenotypes (dbGaP; see go.nature.com/fduced). Likewise, NIH-funded researchers who sequence other HeLa lines will be expected to deposit their data in the dbGaP. We hope that scientists whose work is supported by other funders will do the same.

> *"Non-identifiability is increasingly illusory, owing to technological advances."*

Applications for access to the sequence data will be rapidly reviewed by a newly formed HeLa Genome Data Access working group at the NIH, on which two members of the Lacks family will serve. We believe that this plan reflects the true partnership between the Lacks family and the biomedical-research community. We also ask that all researchers who generate or use genomic data from HeLa cells include in their publications an acknowledgement of the contribution of Lacks and the continued generosity of her family, such as that in Adey and colleagues' paper[3].

Of course, someone could still stitch together a reasonable representation of the HeLa genome from the estimated 1,300 gigabytes of data already in public databases, which have been accumulating over the past 25 years — and the family knows this. The family is also aware that any lab with the right equipment, and non-NIH funds, could derive the full sequence from scratch at any point and post it on a non-NIH website. However, we urge the research community to act responsibly and honour the family's wishes. Downloading the HeLa sequence through controlled access is the right and respectful thing to do.

It is important to note, however, that we are responding to an extraordinary situation here, not setting a precedent for research with previously stored, de-identified specimens. The approach we have developed through working with the Lacks family is unique because HeLa cells were taken and used without consent, and gave rise to the

most widely used human cell line in the world, and because the family members are known by name to millions of people.

The furore around HeLa cells has brought the absence of consent requirements for some biospecimen research to public attention. Under current US federal guidelines, it is still possible to use specimens and to generate whole-genome sequencing data without the knowledge or permission of the person providing the sample, as long as the biospecimen meets the definition of 'de-identified' (see go.nature.com/2jrzvz). The administration of President Barack Obama is undertaking fundamental reforms for the protection of human subjects in research. Among the factors motivating these reforms is the recognition that non-identifiability is increasingly illusory, owing to technological advances, especially in genomics and computing[4–7]. In addition, the relationship between researchers and participants is evolving: seeking permission emphasizes that participants are partners, not just 'subjects'.

In July 2011, the US Department of Health and Human Services issued a notice requesting public comment on how current regulations for protecting participants in research might be revised to be more effective (see go.nature.com/LL6es9). Among other questions, the notice sought comment on whether the department should require consent for future research using samples, identified or not. The notice also sought input on the use of broad consent for unspecified future research use of specimens. The question assumed that specimens that were collected before a change in regulations would be governed by the old rules. On the basis of those public comments, the department is preparing a new proposal.

It is fitting, given the priceless contributions that Henrietta Lacks has made to science and medicine, that her story is catalysing enduring changes in policy. These should afford future generations of research participants the protections and respect that were not in place during Lacks' lifetime. ■ **SEE WORLD VIEW P.123**

Kathy L. Hudson *is deputy director for science, outreach and policy at the National Institutes of Health (NIH) in Bethesda, Maryland.* **Francis S. Collins** *is director of the NIH.*
e-mail: kathy.hudson@nih.gov

1. Landry, J. J. M. *et al. Genes Genomes Genet.* http://dx.doi.org/10.1534/g3.113.005777 (2013).
2. Skloot, R. 'The Immortal Life of Henrietta Lacks, the Sequel' *The New York Times* (23 March 2013).
3. Adey, A. *et al. Nature* **500**, 207–211 (2013).
4. Lin, Z., Owen, A. B., Altman & R. B. *Science* **305**, 183 (2004).
5. Lowrance, W. W. & Collins, F. S. *Science* **317**, 600–602 (2007).
6. Gymrek, M., McGuire, A. L., Golan, D., Halperin, E. & Erlich, Y. *Science* **339**, 321–324 (2013).
7. Rodriguez, L. L., Brooks, L. D., Greenberg, J. H. & Green, E. D. *Science* **339**, 275–276 (2013).

Steel mills and other heavy industries amplify carbon emissions in Inner Mongolia, while the products are consumed in more affluent parts of China.

A low-carbon road map for China

Recycling, renewables and a reinvigorated domestic energy market will allow China to lead the world in low-carbon development, say **Zhu Liu** and colleagues.

China is a major force behind anthropogenic carbon emissions and their mitigation. The world's leading primary energy consumer in 2012, China devoured almost half of all coal produced. The nation accounted for one-quarter of global carbon dioxide emissions in 2011 and 80% of the world's rise in CO_2 emissions since 2008 (ref. 1).

Facing international pressures to curb its CO_2 releases, as well as a tight domestic fossil-energy supply and high levels of air pollution, China has implemented a bold national strategy for energy conservation and emissions mitigation. The country plans to reduce its carbon intensity (CO_2 per unit of gross domestic product, or GDP) to 55–60% of 2005 levels by 2020.

This can be achieved only if China becomes a low-carbon economy. With powerful regulatory control, we believe that the nation's energy appetite could drive the development and use of low-carbon technologies, in which China could become a world leader. We identify the major challenges in such a transition, and propose a five-pronged strategy to get China onto this low-carbon pathway.

First, China must move away from coal and boost recycling and renewable energies. Second, emissions-mitigation indicators, such as energy-efficiency targets, should be set relative to physical output (such as tonnes of steel production) rather than to economic growth. Third, regional energy supply and demand must be balanced. Fourth, energy prices should be linked to market mechanisms rather than set centrally by authorities. And fifth, China must reduce

air pollutants alongside CO_2 emissions.

China has made great progress in cutting carbon emissions in the past decade. In its 11th five-year plan (2006–10), the government set goals to reduce energy intensity (energy consumption per unit of GDP) by 20% on average across all provinces by 2010. Thousands of inefficient power plants and factories were closed to meet the targets[2], saving the equivalent of 750 million tonnes of coal and 1.5 billion tonnes of CO_2 (5% of global CO_2 emissions in 2010).

The government's 12th five-year plan (2011–15) calls for a 16% reduction in energy intensity and for a 17% reduction in carbon intensity. Each region has been allocated mandatory targets. Such cuts would save about 1.4 billion tonnes of coal between 2006 and 2015, reducing CO_2 emissions by more

than 3 billion tonnes (roughly 60% of US emissions in 2010). As a result, air-pollution levels would fall.

LOW-CARBON CHALLENGES

Challenges remain in curbing fossil-energy use and emissions while maintaining economic growth. Although two-thirds of China's provinces met their intensity targets for 2006–10, CO_2 emissions rose nationwide in that period by 50% as the economy grew[3]. Efficiency targets were met by expanding scale of production. When pressure to reach the targets became so great that several provinces instituted blackouts, many factories turned to inefficient diesel power generators, leading to a national diesel shortage in 2010.

Infrastructure construction has been the major driver of China's rapid economic and emissions growth since 2002. As a result, the economy relies on carbon-intensive industries (see 'Economic growth'). From 2005 to 2011, many of China's industries grew faster than its GDP, which rose by 87% (constant price at 2005 value). Thermal power generation grew by 90%, steel production by 135%, cement by 96% and vehicle production by 223% (ref. 4). In 2008, the growth was further exacerbated when China initiated a 4-trillion-renminbi (US$600-billion) economic stimulus plan, of which 85% was for building infrastructure. Today, China accounts for much of the world production of crude steel (45%), cement (60%), primary aluminium (44%), coke (64%) and coal (50%). Almost all of these products are consumed domestically.

China's energy- and emissions-intensity targets are expressed as ratios of energy use or total emissions to GDP. There are thus two ways to achieve the targets: by upgrading equipment and industrial processes to use less energy and to drive down emissions, or by expanding the scale of production, thereby boosting GDP. Both strategies, but especially scale expansion, have contributed to China's improved energy intensity. However, they have resulted in much higher emissions overall. For instance, from 2002 to 2009 China's coal-fired power plants improved their energy intensity by 10%, but because their production capacity more than doubled, total emissions from the sector also doubled[4].

Current emissions targets may exaggerate regional development inequalities by outsourcing the mitigation cost from rich to poor. Some of China's poorest regions rely on carbon-intensive industries, such as cement and steel, and such regions have per capita emissions approaching those of the United States (see 'China emissions'). To support economic development, these regions have been allocated small reduction targets. But most products from these poorer areas are consumed in more affluent regions[5].

For example, in 2010, about 20% of carbon emissions from Inner Mongolia, a poor region in the north of the country, were from the production of electricity that was exported to other provinces. Two-thirds of the region's processed metals, half of its chemical products and 43% of its cement were also sent to more developed areas on the coast.

Beijing and Shanghai import about 70% and 33% of their electricity, respectively[6]. In this way, the cities avoid emitting 50 million tonnes and 38 million tonnes of CO_2, respectively. If each region's emissions, and those embodied in traded products, were allocated to final consumers, Inner Mongolia would have exceeded its energy-intensity reduction commitments by around 40% in 2010, whereas Shanghai and Beijing would have failed to achieve theirs.

The immaturity of China's electricity grid results in power shortages and inefficiencies. Many new power plants remain unconnected to the national grid, owing to poor coordination between local governments and the State Grid Corporation of China, which builds and manages all grids in the country. One-third of Inner Mongolia's power capacity (100 billion kilowatt hours, or kWh) remains unused each year for this reason.

Because the prices of fuels such as coal are set by and fluctuate with the market, but the cost of electricity consumption is fixed by the central government, power plants slow their production rate when fuel prices are high. This is inefficient and causes blackouts. In 2010, China produced only about half of its capacity of 6,220 billion kWh.

LOW-CARBON LEAPFROGGING

China's economy can be decarbonized only by reducing fossil-energy demand and emissions together. Recycling construction materials could reduce total energy intensity by as much as 90%. China recycled about 70 million tonnes of steel scrap in 2008, and scrap has the potential to replace 80% of iron ore as a resource for primary steel production by 2050 (ref. 7). Schemes encouraging 'urban

mining' of scrap and exchanging by-products among regional factories are needed.

China leads the world in renewable energy, having invested $68 billion in 2012 — more than one-fifth of the global total for that year. The country's installed renewable capacity of 300 gigawatts (GW) in 2011 was already twice the US capacity for the same year (146 GW). China's wind turbines and hydropower stations were the world's most productive in 2011, generating 70 billion kWh and 720 billion kWh, respectively.

Yet China is producing more renewable technologies than it can use. In 2012, the country's manufacturing capacity for solar photovoltaic cells reached 40 GW. But it produced only 23 GW, which accounts for 60% of world's total annual production (37 GW). Less than 10% of these domestically produced photovoltaics were installed in the country — in part owing to the lack of grid connections. The rest were exported. Further investments are urgently needed to extend the electricity grid to remote rural regions, where solar power is a solution for electrification, and to explore other markets for the surplus supply of panels.
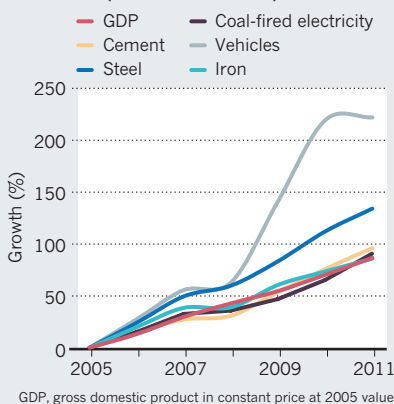
There is great potential for low-carbon energy technologies in China. Wind power alone could meet the entire projected increase in electricity demand up to 2030. Introducing 640 GW of wind capacity (costing about $900 billion) over the next 20 years will reduce carbon emissions by 30% in the period[8]. Using China's waste gutter oil (13.7 million tonnes in 2010), which is refined from discarded cooking oil, for biomass fuel could reduce CO_2 emissions by some 90 million tonnes per year[9]. Such a reduction would be equivalent to 15% of the total emissions reduction from 1990 to 2008 by the 39 industrialized countries under Annex B of the Kyoto Protocol — the international treaty that commits parties to cutting greenhouse-gas emissions.

In the meantime, cleaner, non-renewable options, such as natural gas and nuclear power, could provide a buffer during the low-carbon transition. Through further exploitation of coal-bed methane and by improving connectivity of domestic and international gas pipelines, natural-gas consumption in China could swell to 250 billion cubic metres by 2020, with double-digit growth from 2010 to 2020. Switching from coal to gas would simultaneously reduce air pollutants, such as sulphur dioxide and nitrogen oxides, as well as CO_2.
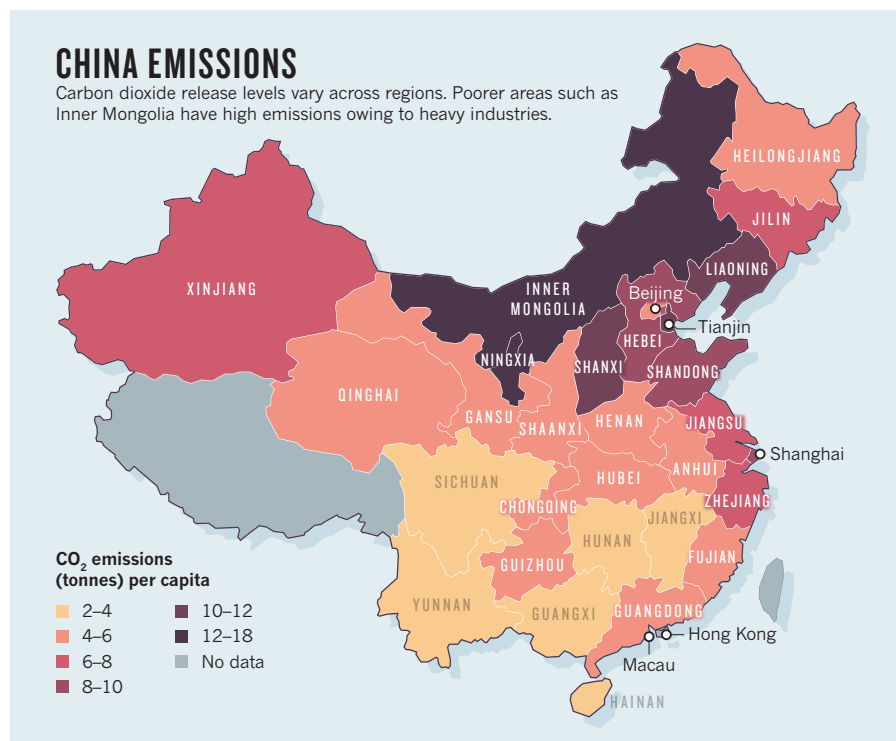
With 17 reactors in operation, nuclear power currently accounts for 1% (13 GW) of China's electricity production capacity. Another 28 nuclear plants are under construction and more technologically advanced reactors are planned. The total nuclear capacity is set to rise to 80 GW by 2020, 200 GW by 2030, and 400 GW by 2050.

ECONOMIC GROWTH

China's manufacturing output has risen rapidly since 2005 (relative levels shown).

- GDP
- Cement
- Steel
- Coal-fired electricity
- Vehicles
- Iron

GDP, gross domestic product in constant price at 2005 value

SOURCE: REF. 4

CHINA EMISSIONS

Carbon dioxide release levels vary across regions. Poorer areas such as Inner Mongolia have high emissions owing to heavy industries.

CO$_2$ emissions (tonnes) per capita

- 2–4
- 4–6
- 6–8
- 8–10
- 10–12
- 12–18
- No data

To measure and promote progress, energy and emissions targets must be considered separately to economic performance. Physical intensity indicators, such as emissions per unit production of steel, should be used rather than relative economic intensity indicators, such as emissions per unit of GDP. China has set a coal-consumption cap of about 3.9 billion tonnes by 2015. Similarly, a national carbon-emissions cap should be introduced, leveraged by energy taxes and allowances.

A carbon budget that considers both emissions and offsets from carbon sinks should be introduced for CO$_2$ inventories. This would encourage forest planting and waste management through carbon-credit schemes. The carbon sink created by China's enhanced afforestation and territorial ecosystem might have taken up as much as 15% of China's fossil-fuel CO$_2$ from 2002 to 2007 (ref. 10).

Regional compensation mechanisms would accelerate technology transfer between Chinese provinces. Targets for industrial sectors, rather than regions, would lessen geographic disparities. Regions should count emissions according to electricity consumption rather than by production. Mitigation responsibilities should be required throughout the supply chain of energy-intensive enterprises, from company headquarters in rich areas to factories in poor ones. Rigorous environmental standards need to be strictly applied.

China's first domestic carbon market for permits to discharge CO$_2$ opened in June this year in Shenzhen, with permits for 20,000 tonnes of CO$_2$ traded in the first day. Full marketization with moderate administrative

intervention in China's fuel consumption will be crucial for the success of its domestic cap-and-trade scheme, which is currently being tested in seven provinces and municipalities, affecting an annual 1.5 billion tonnes of CO$_2$ emissions. The system will be implemented nationwide in 2016.

Through the Kyoto Protocol, China is hosting more than 2,000 projects (half of the world total) under the Clean Development Mechanism, which can offset about 600 million tonnes of CO$_2$ from China. The domestic cap-and-trade system is expected to cover 1 billion tonnes of CO$_2$ per year by 2015 (about 10% of China's total CO$_2$ emissions) and generate billions of US dollars of government revenue.

A successful cap-and-trade scheme requires reliable carbon data, a transparent carbon market and fair credit allocations. China's central government should compile and verify the emissions inventories; coordinate, monitor and report on the market measures; define the reduction baselines and certify emissions-reduction credits.

The wealthiest 5% of China's population accounts for 25% of electricity use. The government should pioneer other economic mechanisms, such as a carbon tax, that target consumption. As a first step, China has implemented a household 'incline block tariff' since July 2012, whereby electricity purchased beyond a certain limit has a higher price. Similar taxes should be introduced for other consumables (such as vehicle fuels) and the revenue used to subsidize low-carbon products (such as electric cars), renewable-energy development and

energy infrastructure construction.

Because CO$_2$ emissions and air pollution largely stem from fossil-fuel use, regional control strategies for both should be integrated. China implemented regulations in June this year and plans to invest 1.7 trillion renminbi between 2013 and 2017 to limit urban air pollution, including particulates and ozone. These regulations suggest phasing out inefficient industrial boilers, limiting the expansion of emissions-intensive industries and enhancing regulations and market stimuli for green-energy development, which would accelerate the development of energy saving and environmental protection industries.

China's energy-management system is being restructured under new government leadership. The streamlining associated with the annexing of the State Electricity Regulatory Commission by the National Energy Administration should help to harmonize energy prices and policies across central and local governments and enterprises. A high-level governmental organization (such as the State Council) is also required to coordinate energy policies across agencies, such as the Ministry of Environmental Protection, National Development and Reform Commission, and provincial governments.

By tackling these challenges, we believe that China can lead the global climate-change mitigation movement and create a pathway towards sustainable, low-carbon development. ∎

Zhu Liu *is a fellow in the Sustainability Science Program at Harvard University, Cambridge, Massachusetts; and at the Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang, China.* **Dabo Guan** *is associate professor of environmental economics at the University of Leeds, UK.* **Douglas Crawford-Brown** *is director of the Cambridge Centre for Climate Change Mitigation Research, University of Cambridge, UK.* **Qiang Zhang** *is professor of atmospheric chemistry at Tsinghua University, Beijing.* **Kebin He** *is dean of the School of Environment at Tsinghua University, Beijing.* **Jianguo Liu** *is director of the Center for Systems Integration and Sustainability at Michigan State University, East Lansing.*
e-mail: d.guan@leeds.ac.uk

1. Peters, G. P. et al. Nature Clim. Change **2**, 2–4 (2012).
2. Liu, Z. et al. Energy Policy **49**, 751–758 (2012).
3. Liu, Z. et al. Energy **45**, 1059–1068 (2012).
4. National Bureau of Statistics of China. China Statistical Yearbook 2012 (2012).
5. Feng, K. et al. Proc. Natl Acad. Sci. USA **110**, 11654–11659 (2013).
6. Liu, Z. et al. Energy **37**, 245–254 (2012).
7. Pauliuk, S., Wang, T. & Müller, D. B. Environ. Sci. Technol. **46**, 148–154 (2011).
8. McElroy, M. B. et al. Science **325**, 1378–1380 (2009).
9. Liang, S., Liu, Z., Xu, M. & Zhang, T. Bioresource Technol. **131**, 139–145 (2013).
10. Pan, Y. et al. Science **333**, 988–993 (2011).

DON BAYLEY/GETTY

A rogue banana peel creates the potential for a bout of *Schadenfreude*.

SOCIAL PSYCHOLOGY

The gloat factor

Dan Jones mulls over a study of why we enjoy the misfortunes of others.

In the summer of 2012, best-selling science writer Jonah Lehrer suffered a dramatic and public fall from grace. It became apparent that Lehrer had recycled the words of at least one other writer and had even invented quotes from Bob Dylan in his most recent book, *Imagine: How Creativity Works* (Houghton Mifflin Harcourt, 2012).

Lehrer lost his new job at *The New Yorker*, his publishers pulled his books and writers across the Twitterverse and blogosphere referred to him with scorn. Yet beneath the righteous indignation of his many critics lurked a sense of pleasure in seeing this young, hip, successful author cut down to size.

Why do the misfortunes of others give us a lift? This is the question explored by social psychologist Richard Smith in *The Joy of Pain*, a breezy but serious exploration of the phenomenon. Smith's answer is that *Schadenfreude* — an emotion as ignoble as envy or spite, from the German for harm (*Schaden*) and joy (*Freude*) — pays psychological dividends by enabling us to feel better about ourselves, and our social worth and rank, through "downward comparison" with others.

For Smith, *Schadenfreude* is grounded in our evolved social psychology. Life is a competitive game, with winners and losers in the search for status, mates and much else; and as far as natural selection is concerned, what matters is not your absolute level of success, but how much better or worse you are doing relative to everyone else. So humans are keenly aware of how our attributes, skills and successes stack up against the game's other players and when we come off badly in these social comparisons, our self-esteem takes a hit.

Likewise, seeing those who are above us in social rank take a fall boosts our own relative standing and makes us feel good. Smith discusses experimental studies by social psychologist Wilco van Dijk and his colleagues that show that when people's self-esteem is challenged (by being given bad but false feedback on a test they've taken), they are more likely to take pleasure in hearing about a successful person coming undone. Similarly, the researchers showed that people

The Joy of Pain: Schadenfreude and the Dark Side of Human Nature
RICHARD H. SMITH
Oxford University Press (USA): 2013.

with pre-existing low self-esteem who take part in studies are prone to *Schadenfreude*.

Smith describes many routes to *Schadenfreude*, from the relatively passive — such as comparing ourselves to people who are down and out, and exaggerating negative qualities (or dismissing positive qualities) of those more successful than us — to actively bringing about the misfortune of people we envy.

He also notes a number of factors that prime us for *Schadenfreude*. It is likely to bubble up when we think that someone's misfortune is a case of just deserts, and never more so than when the person is revealed to be a hypocrite — such as the many evangelical preachers exposed as indulging in the behaviours they condemn in others.

Envy also amps up *Schadenfreude*, as memorably articulated by writer Clive James's poem 'The book of my enemy has been remaindered [and I am pleased]' (see go.nature.com/gr5kdf). Conflicts and competition between groups are likely to bring out *Schadenfreude*; think of sport, in which pleasure in the misfortune of rivals is socially acceptable. Psychologist Charles Hoogland and his colleagues, for example, have shown that committed basketball fans are pleased when rival team members suffer even severe injuries. And brain-imaging studies by psychologist Susan Fiske and others revealed that when fans of baseball team the Boston Red Sox witness their team beating arch-rivals the New York Yankees (or vice versa), their brains show more activation of reward systems than when their team beats a more neutral opponent, underscoring the importance of competitive drive in *Schadenfreude*. Politics is yet another rich seam: recall the celebratory parties that erupted when former UK prime minister Margaret Thatcher died earlier this year.

Schadenfreude fuelled by a combination of resentment and our divisive tendency to form exclusive, competitive groups can be especially potent, bringing out the darkest sides of human nature and leading people to actively engineer misfortune in other groups. Smith suggests that such a process plausibly had a role in Nazi propaganda, which was explicitly designed to arouse resentment, envy and enmity towards Jewish people, and so to offer a specious justification for their subsequent extreme mistreatment and incalculable suffering.

Smith's portrait of this complex response combines experimental studies with many well-chosen examples drawn from political scandals, biographies, reality-television shows, literature, sitcoms, cartoons and the observations of comedians and satirists. *The Joy of Pain* is a real joy to read — and completely painless. ∎

Dan Jones *is a freelance science writer based in Brighton, UK.*
e-mail: dan.jones@multipledrafts.com

Food is handed out at a hospital in the Central African Republic.

DEVELOPMENT

Starved for solutions

Calestous Juma weighs up a call for a revolution to end world hunger.

Some 870 million people suffer from chronic undernourishment, despite humanity's best efforts to improve agricultural productivity, create markets and boost nutrition. In *Betting on Famine*, sociologist Jean Ziegler sets out to provide a human rights-based approach to addressing world hunger. The book is a sweeping indictment of global injustice and provides ample facts and figures. "The destruction, every year, of tens of millions of men, women, and children from hunger is the greatest scandal of our era," says Ziegler, who was United Nations (UN) special rapporteur on the right to food from 2000 to 2008.

His main thesis, which is in no way innovative, is that the world is capable of feeding 12 billion people — 5 billion more than now exist. The main obstacle, in his view, is global inequality and corporate control of the food system. The solution, he says, is to return to the fundamental principles of the right to food, defined by the UN as having "regular, permanent and unrestricted access, either directly or by means of financial purchases, to quantitatively and qualitatively adequate

and sufficient food corresponding to the cultural traditions of the people".

Ziegler argues that access to food has been one of the most flouted human rights in history. He attributes much of the reason for this to the dominance of the private sector and an unfair global trading system, underpinned by what he sees as neo-liberal dogma, such as the perceived benefits of privatizing public enterprises. He argues that nothing short of a revolution is needed to curb corruption among leaders in emerging nations most affected by famine, promote popular resistance among social movements around the world, and make the right to food a policy priority in parliamentary and other governance bodies.

Betting on Famine disappoints for many reasons, one being that it says nothing new. Classics such as Susan George's *How the*

Betting on Famine: Why the World Still Goes Hungry
JEAN ZIEGLER
The New Press: 2013.

Other Half Dies (1976) have provided more incisive assessments of why famine has persisted despite increases in food production. Ziegler admits that much of what needs to be done has already been outlined in numerous UN documents. Furthermore, his book is primarily a diatribe against those in power; it offers little by way of example or inspiration on how to solve world hunger. Appealing to revolution is possibly the easiest of intellectual expeditions. Executing the task is much more complex and requires the involvement of the same corporations and governments that the book incessantly admonishes.

There is an equally revolutionary alternative that Ziegler does not acknowledge: empowering the poor by building their capacity to address hunger through improved agricultural practices, training of farmers, better infrastructure and access to markets. Following the 1974 coup in Ethiopia, for instance, Marxist leaders embarked on a peasant revolution aimed at overthrowing landowners in the hope that this would lead to the modernization of agriculture. It did not work. But now the country's government focuses on

promoting and expanding cooperation between farmers and the same corporations that Ziegler wishes to send to the gallows. Partly because of improvements in agricultural production, Ethiopia's economy has registered an average growth of 8% per year in the past decade.

Rights cannot be wished into existence. They need institutions to become realities. In 2010, Ethiopia created the Agricultural Transformation Agency (ATA), mirroring elements of Brazil's Agricultural Research Corporation (Embrapa), which has helped to bring technical support and credit to farmers. The ATA focuses on empowering farmers to become more entrepreneurial by helping them to improve productivity and participate in local and global markets. Ethiopia is also now a member of the Grow Africa consortium, which includes private enterprises, the African Union and the World Economic Forum, and has pledged to invest more than US$3.5 billion in African agriculture. China, India and Brazil, among other countries, are also actively tackling hunger with more inclusive approaches, accommodating all major players including private corporations.

> *"Rights cannot be wished into existence. They need institutions to become realities."*

Ziegler rightly emphasizes the role of farmers, but fails to note how technical training can strengthen their political influence. Innovations such as the US land-grant university model, formalized 150 years ago to bring agricultural research, teaching and extension under one roof, played a key part in educating US farmers. The Green Revolution that helped countries such as India and Mexico to avert major famines relied heavily on scientific research, participation of the private sector and the upgrading of farmers' skills.

The right to food will continue to be a major global challenge as pressure on natural resources increases. But solutions will not come from traditional appeals for popular uprisings. They will come from increased inclusivity in partnerships, involving rather than punishing private corporations. To feed the hungry, the world needs new approaches that expand the practical use of human creativity, not more pleas for hollow revolutions. ∎

Calestous Juma *is professor of the practice of international development at Harvard Kennedy School in Cambridge, Massachusetts, and author of* The New Harvest: Agricultural Innovation in Africa.
e-mail: calestous_juma@harvard.edu

SCIENCE AND RELIGION

Godless chronicles

Glenn Branch goes for a dip in the antitheologic.

When the producers of the 2008 creationist film *Expelled* asked to interview PZ Myers, they misrepresented the nature of their project and the purpose of the interview. By the time the film screened, Myers knew that his brief clip would portray the scientific establishment as dogmatically suppressing dissent from evolution. Ironically, when he arrived at the cinema, he was excluded — unlike a colleague who was also interviewed for the film. Myers repaired to a nearby computer store to post a hilarious account of his expulsion on his widely read blog, Pharyngula. *Expelled* was quickly showered with unwelcome publicity as a result.

Myers is a developmental biologist, who named his blog after the pharyngula stage of embryonic development — both, he jokes, are notable for the appearance of brain and jaw. Pharyngula is a freewheeling mix of explanations of developmental biology, denunciations of creationism, commentary on politics, feuds with critics and rivals, and the sort of in-jokes and recurrent features that typify the blogosphere, enlivened by a raucous chorus of commenters. Its slogan is: "Evolution, development, and random biological ejaculations from a godless liberal".

A major theme of Pharyngula, and Myers's first book, *The Happy Atheist*, is what he views as the incompatibility of science and religion. In addition to excoriating various absurdities and atrocities that he associates with faith, such as the bad science deployed by anti-abortion zealots, Myers repeatedly asserts that science and religion are necessarily in conflict: "One is a method of analysis and experiment; the other is pretense and lies." He is fierce with regard to the proponents of old-fashioned creation science ("trying to get their Old Testament superhero to adhere to the rules of physics, chemistry, biology, and ordinary common sense") and the adherents of newfangled intelligent design (who "hide the bearded old sky god from the public eye"). He also castigates scientists who accept evolution while retaining their faith.

Whatever Myers's target, his weapons are taken from the arsenal of ridicule. He is in good company — writers such as Jonathan Swift and George Orwell spring to mind. Myers's prose, although serviceable, isn't quite in the same class, but sometimes reaches lyrical heights. Explaining his decision to bury, rather than burn, unwanted books of scripture sent for his spiritual
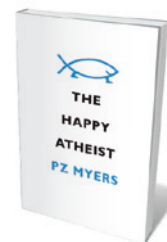
The Happy Atheist
PZ MYERS
Pantheon: 2013.

instruction, he exults "as nematodes writhe over the surfaces, etching the words with slime and replacing the follies of dead men with the wisdom of worms". Myers's favourite weapon is the extended metaphor, deployed to expose his targets as arbitrary and absurd. He wields it adroitly, comparing religious diversity to hat variety and theologians to courtiers fawning over the Emperor's new clothes. These conceits are often amusing and occasionally instructive, but the tactic is cheap.

Whether infuriating or invigorating, ridicule is no substitute for a considered critique, and Myers often fails to do justice to his targets. For example, his analysis of the idea that God guides evolution by acting undetectably at the quantum level, if amusing, is a popular rather than a scholarly treatment, and incorporates value judgements that are unsupportable by science. Myers might respond that his targets are too ridiculous to warrant anything more serious, but such a response presupposes, rather than compels, agreement.

The chief problem with *The Happy Atheist*, however, is that it seems to break no new ground. By my count, Pharyngula posts provide the basis for at least 26 of the 38 essays and 5 more are adapted from a talk he gave in 2010.

Admirers and detractors alike will be disappointed by the book as a missed opportunity for Myers to refine, systematize and extend his thoughts on science and religion. It is not comparable with Jason Rosenhouse's *Among the Creationists* (Oxford University Press, 2012), Steve Stewart-Williams's *Darwin, God and the Meaning of Life* (Cambridge University Press, 2010), or that 'summa antitheologica' of our day, *The God Delusion* (Houghton Mifflin Harcourt, 2006) by Richard Dawkins. It was Dawkins, by the way, who was admitted to the screening of *Expelled* when Myers was excluded. Was Voltaire's prayer, "O Lord, make our enemies ridiculous," ever better answered? ∎

Glenn Branch *is deputy director at the National Center for Science Education in Berkeley, California.*
e-mail: branch@ncse.com

Correspondence

Gain-of-function experiments on H7N9

Since the end of March 2013, avian influenza A viruses of the H7N9 subtype have caused more than 130 human cases of infection in China, many of which were severe, resulting in 43 fatalities. Although this A(H7N9) outbreak is now under control, the virus (or one with similar properties) could re-emerge as winter approaches.

To better assess the pandemic threat posed by A(H7N9) viruses, investigators from the NIAID Centers of Excellence in Influenza Research and Surveillance and other expert laboratories in China and elsewhere have characterized the wild-type avian A(H7N9) viruses in terms of host range, virulence and transmission, and are evaluating the effectiveness of antiviral drugs and vaccine candidates. However, to fully assess the potential risk associated with these novel viruses, there is a need for further research, including experiments that may be classified as 'gain of function' (GOF).

Here we outline the aspects of the current situation that most urgently require additional research, our proposed studies, and risk-mitigation strategies.

The A(H7N9) virus haemagglutinin protein has several motifs that are characteristic of mammalian-adapted and human influenza viruses, including mutations that confer human-type receptor binding and enhanced virus replication in mammals. The pandemic risk rises exponentially should these viruses acquire the ability to transmit readily among humans.

Reports indicate that several A(H7N9) viruses from patients who were undergoing antiviral treatment acquired resistance to the primary medical countermeasure — neuraminidase inhibitors (such as oseltamivir, peramivir and zanamivir). Acquisition of

resistance to these inhibitors by A(H7N9) viruses could increase the risk of serious outcomes of A(H7N9) virus infections.

The haemagglutinin proteins of A(H7N9) viruses have a cleavage site that is consistent with a low-pathogenic phenotype in birds. In the past, highly pathogenic H7 variants (with basic amino-acid insertions at the cleavage site that enable the spread of the virus to internal organs) have emerged from populations of low-pathogenic strains circulating in domestic gallinaceous poultry.

Normally, epidemiological studies and characterization of viruses from field isolates are used to inform policy decisions regarding public-health responses to a potential pandemic. However, classical epidemiological tracking does not give public-health authorities the time they need to mount an effective response to mitigate the effects of a pandemic

virus. To provide information that can assist surveillance activities — thus enabling appropriate public-health preparations to be initiated before a pandemic — experiments that may result in GOF are critical.

Therefore, after review and approval, we propose to perform experiments that may result in GOF (see 'Proposed gain-of-function experiments').

All experiments proposed by influenza investigators are subject to review by institutional biosafety committees. The committees include experts in the fields of infectious disease, immunology, biosafety, molecular biology and public health; also, members of the public represent views from outside the research community. Risk-mitigation plans for working with potentially dangerous influenza viruses, including the 1918 virus and highly pathogenic avian H5N1 viruses, will be applied to conduct

GOF experiments with A(H7N9) viruses (see Supplementary Information at go.nature.com/fstdy1). Additional reviews may be required by the funding agencies for proposed studies of A(H7N9) viruses.

The recent H5N1 virus-transmission controversy focused on the balance of risks and benefits of conducting research that proved the ability of the H5N1 virus to become transmissible in mammals (see www.nature.com/mutantflu). These findings demonstrated the pandemic potential of H5N1 viruses and reinforced the need for continued optimization of pandemic-preparedness measures. Key mutations associated with adaptation to mammals, included in an annotated inventory for mutations in H5N1 viruses developed by the US Centers for Disease Control and Prevention, were identified in human isolates of A(H7N9) viruses. Scientific evidence of the pandemic threat posed by A(H7N9) viruses, based on H5N1 GOF studies, factored in risk assessments by public-health officials in China, the United States and other countries.

Since the H5 transmission papers were published, follow-up scientific studies have contributed to our understanding of host adaptation by influenza viruses, the development of vaccines and therapeutics, and improved surveillance.

Finally, a benefit of the H5N1 controversy has been the increased dialogue regarding laboratory biosafety and dual-use research. The World Health Organization issued laboratory biosafety guidelines for conducting research on H5N1 transmission and, in the United States, additional oversight policies and risk-mitigation practices have been put in place or proposed. Some journals now encourage authors to include biosafety and biosecurity descriptions in their papers, thereby raising the awareness of

---

H7N9 INFLUENZA RESEARCH
Proposed gain-of-function experiments

● **Immunogenicity.** To develop more effective vaccines and determine whether genetic changes that confer altered virulence, host range or transmissibility also change antigenicity.

● **Adaptation.** To assist with risk assessment of the pandemic potential of field strains and evaluate the potential of A(H7N9) viruses to become better adapted to mammals, including determining the ability of these viruses to reassort with other circulating influenza strains.

● **Drug resistance.** To assess the potential for drug resistance to emerge in circulating viruses, evaluate the genetic stability of mutations conferring drug resistance, and evaluate the efficacy of combination therapy with

antiviral therapeutics. Also, to determine whether A(H7N9) viruses could become resistant to available antiviral drugs, and to identify potential resistance mutations that should be monitored during antiviral treatment.

● **Transmission.** To assess the pandemic potential of circulating strains and perform transmission studies to identify mutations and gene combinations that confer enhanced transmissibility in mammalian models (such as ferrets and guinea pigs).

● **Pathogenicity.** To aid risk assessment and identify mechanisms, including reassortment and changes to the haemagglutinin cleavage site, that would enable circulating A(H7N9) viruses to become more pathogenic.

researchers intending to replicate experiments.

The risk of a pandemic caused by an avian influenza virus exists in nature. As members of the influenza research community, we believe that the avian A(H7N9) virus outbreak requires focused fundamental and applied research conducted by responsible investigators with appropriate facilities and risk-mitigation plans in place. To answer key questions important to public health, research that may result in GOF is necessary and should be done.

Ron A. M. Fouchier* *Erasmus Medical Center, Rotterdam, the Netherlands.*
r.fouchier@erasmusmc.nl
Yoshihiro Kawaoka* *University of Wisconsin-Madison, Wisconsin, USA.*
kawaokay@svm.vetmed.wisc.edu
*On behalf of 22 co-authors (see go.nature.com/fstdy1 for full list).

Extra oversight for H7N9 experiments

The US Department of Health and Human Services (HHS) announces a new review process for certain gain-of-function (GOF) experiments on the avian influenza A (H7N9) virus, some of which are proposed this week by influenza scientists (R. A. M. Fouchier *et al. Nature* **500,** 150–151; 2013).

Specifically, before being undertaken using HHS funds, any experiments that are reasonably anticipated to generate H7N9 viruses with increased transmissibility between mammals by respiratory droplets will undergo an additional level of review by the HHS.

The HHS review will consider the acceptability of these experiments in light of potential scientific and public-health benefits as well as biosafety and biosecurity risks, and will identify any additional risk-mitigation measures needed. The review will be carried out by a standing multidisciplinary panel of federal experts with backgrounds in public health, medicine, security, science policy, global health, risk assessment, US law and ethics.

This approach, similar to that for certain H5N1 influenza virus experiments (see go.nature.com/vpmplf), allows the HHS to focus special oversight efforts on experiments of concern while allowing routine characterization and other fundamental research to proceed rapidly, thereby enabling a robust public-health response.

GOF studies can provide important insights into how the A(H7N9) virus adapts to mammalian hosts, causes disease and spreads to other hosts, but they may also pose biosafety and biosecurity risks. To ensure that research involving H7N9 virus is conducted safely and securely, the US Centers for Disease Control and Prevention recently re-examined the requisite biosafety conditions for conducting experiments involving H7N9 and, in June 2013, issued interim risk-assessment and biosafety-level recommendations (see go.nature.com/gknn9a).

Harold W. Jaffe *Centers for Disease Control and Prevention, Atlanta, Georgia, USA.*
Amy P. Patterson *National Institutes of Health, Bethesda, Maryland, USA.*
pattersa@od.nih.gov
Nicole Lurie *Department of Health and Human Services, Washington DC, USA.*

Follow Obama's lead and take a pay cut

In considering the impact of the US budget sequester on science (see, for example, *Nature* **499,** 147–148; 2013), I see no mention of salary reductions. A 5% reduction in the salaries of federally supported science staff, including administrative and agency personnel, would significantly reduce the need to cut science programmes.

US scientists are not paupers: NASA scientists, for instance, are paid up to US$160,000 a year, with generous fringe benefits. Plenty of professors at leading US universities make much more. A 5% cut to 12-month and summer salaries would not leave anyone destitute.

Such a reduction would be much more effective than any presentations to Congress in showing that scientists care about their projects and are willing to share the pain of bringing US federal expenditure under control.

I doubt that any scientist would refuse to accept a grant offered on the proviso that the salary rate be reduced by 5%, if the alternative were no grant at all. President Barack Obama took a pay cut to show the way.

Let's follow his lead.
Peter Foukal *Nahant, Massachusetts, USA.*
pvfoukal@comcast.net

Three reasons for eco-label failure

The fisheries industry promotes third-party eco-labels that signify sustainability, similar to those used in forestry and tourism (see *Nature* http://doi.org/nb5; 2013). In my view, these fail for three reasons.

First, consumers care strongly that labels for health and quality standards are accurate because they affect individuals, but care much less about eco-labels because their effects are spread across society.

Second, industries tend to use weak eco-labels in political games to avoid strong regulation.

Third, ineffective eco-labels closely mimic accurate ones. Because there are no adverse consequences for consumers who cannot tell them apart, a high proportion of mimics persists.

Eco-labels are thus no substitute for eco-laws.
Ralf Buckley *Griffith University, Gold Coast, Queensland, Australia.*
r.buckley@griffith.edu.au

NEWS & VIEWS

Accurate maps of visual circuitry

Such is the brain's complexity that even small neural circuits contain hundreds of neurons making thousands of connections. Connectivity and optical analyses provide close-up views of two such circuits. SEE ARTICLES P.168, P.175 & LETTER P.212

RICHARD H. MASLAND

Understanding the biological machinery from which perception, action and thinking are built is not an easy undertaking. A big difficulty is that neuroscientists must deal with a problem of spatial scale — one in which components range from nanometre-sized synaptic junctions between neurons to centimetre-long connections between brain regions — and study these scales simultaneously. Three papers[1–3] in this issue attack these problems of scale. Two of them (by Helmstaedter et al.[1] and Takemura et al.[2]) use computational techniques to expand the field of neurons that can be encompassed in a high-resolution view. The third study (by Maisak et al.[3]) combines genetic and optical methods to record the activity of neurons that until now have been impossible to monitor owing to their small size. All three take as a model the retina, the first image-processing element in the chain that leads to visual perception.

The mammalian retina contains more than 60 different types of neuron, each of which has a distinct morphology and carries out a different function[4]. Within the retina, photoreceptor cells sense light, and their output is processed by amacrine, horizontal and bipolar cells. Downstream, roughly 20 different types of retinal ganglion cell transmit the final coded signal — 20 different representations of the visual input — to the brain. Unsurprisingly, therefore, sorting out neuronal connectivity in the retina has been a daunting task. Helmstaedter et al. (page 168) now report a connectome (a list of all synaptic connections) for an inner layer of the mouse retina. They achieve this by serial tissue sectioning and electron microscopy, followed by digital reconstruction of cells within the virtual three-dimensional solid that results.

The analysis reveals patterns of connections that could account for the stimulus selectivity of two types of ganglion cell. More fundamentally, the reconstruction, which contains 950 neurons (Fig. 1a), allows definitive classification of the types of bipolar cell. With only a slight refinement, the new classification matches extremely well with the existing understanding of these cells[5], which was
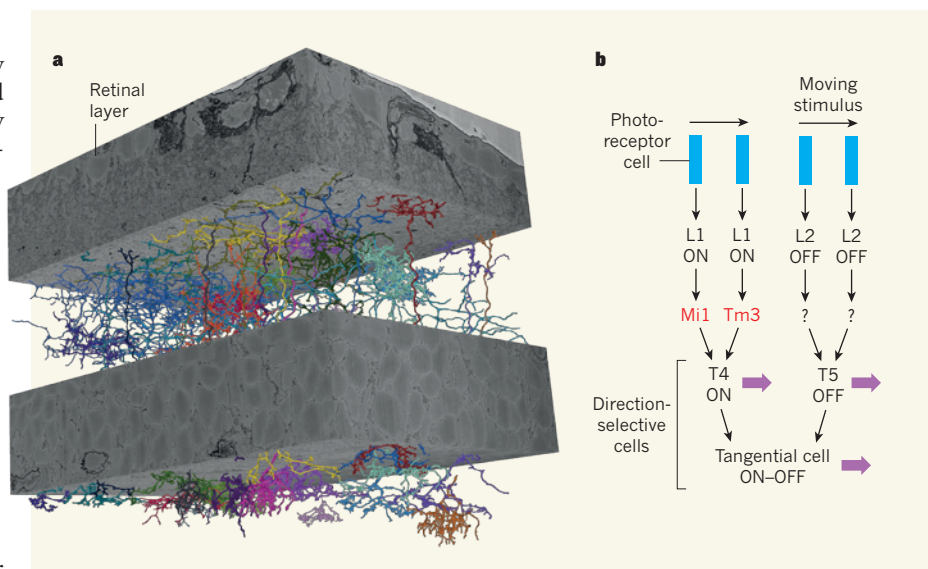
Figure 1 | The mechanism of motion discrimination in the visual system[1–3]. a, Reconstruction of 24 out of 950 mouse neurons between two retinal layers, based on an electron microscopy data set. **b**, Photoreceptor cells slightly separated in space mediate inputs to Mi1 and Tm3 cells, through intermediary L1 and L2 cells. Their outputs converge on the T4 cells, which, because of the spatial separation of the inputs, can discriminate movements occurring in different directions. A similar mechanism is thought to occur in T5 cells, although the intermediary cells are unknown. T4 and T5 cells are selectively responsive to light edges (ON) or dark edges (OFF), respectively. Thus, at the level of these cells, the visual input is broken down into eight separate components, each representing ON or OFF activity and one of four directions of motion (only forward motion is shown; purple arrows). This information is then recombined by the tangential cells, each of which is sensitive to one of four cardinal directions, and to both ON and OFF edges. For anatomically correct details of the diagram, see Figure 4 of ref. 2.

based largely on the identification of molecular markers using light microscopy. Helmstaedter and colleagues' work, however, adds value by providing enormously more precise descriptions of bipolar-cell structure and by effectively acting as a positive control, increasing confidence that analysis of the amacrine and ganglion cell types, which have resisted classification by previous techniques, will be equally definitive. And this is just the beginning: once these cell types are classified, the same basic methods should allow the synaptic connections among them to be deciphered.

Takemura et al. (page 175) and Maisak and co-workers (page 212) report progress on a classic problem of neural computation — the detection of visual motion. Their test system is the eye of the fruitfly, an animal that must rapidly navigate during flight and that is especially effective at dodging predators (doubters

are invited to swat one). It is easy to make simple models of motion detection[6,7], but pinning the mechanism to precise neural events has been much harder. Whereas photoreceptor cells cannot detect direction, downstream neurons called tangential cells are robustly tuned to the direction of movement. Somewhere in between lies the neural mechanism that creates the directional discrimination, but the crucial neurons, called T4 and T5, are too small for ordinary electrical recording. Maisak et al. got around this difficulty by recording activity optically, using an indicator protein introduced into the cells by genetic techniques.

The authors demonstrate that T4 and T5 detect visual movement, with subsets of each being selective for one of four cardinal directions: upward, downward, front to back, and back to front. Furthermore, these cells are sensitive to opposite visual contrasts — T4 cells

respond to light ON and so are sensitive to light edges, whereas T5 cells respond to light OFF and are sensitive to dark edges. The authors' genetic-knockout experiments not only confirm this optical observation but also show that T4 and T5 are the sole pathways mediating these functions, with no other cells being able to step in and carry the message. Thus, a fly initially breaks down moving visual inputs into a total of eight components: bright edges moving up, down, forward or backward, and dark edges moving along the same four axes.

But how do T4 and T5 actually detect the direction of movement? Takemura and collaborators' fly connectome suggests a place to look for the answer. They show that just upstream of T4 lies a pair of neurons, termed Mi1 and Tm3, which report on narrowly separated points in visual space. Because of that separation, the pair could provide the inputs that T4 needs to discriminate direction (Fig. 1b). Using Maisak and colleagues' recording technology, it may be possible to optically record from Mi1 and Tm3. If all goes well, this could bring the 50-year search for the mechanism of direction selectivity to an end.

The connectomic approach has now proved its importance for studying fly eyes and mouse retinas, but sceptics will still doubt that we can make the jump from these miniature neuronal circuits to 'real brains'; the intrinsic circuits of the cerebral cortex are some ten times larger than those of the retina, and this spatial scale is dwarfed again by the distances that connect different brain regions.

One obstacle is the need for very large tissue sections. Another difficulty is that of image segmentation, which is required for tracing thin neuronal processes through the thicket of neighbours in serial sections. Because digital solutions have failed, the task is currently assigned to large teams of human observers, but this will be impractical on larger spatial scales. Improvements in fixation and staining might make the processes more easily discriminable, and digital technology may yet save the day, because in principle any task that can be done by human observers can be done by a computer. Tracing processes is essentially a problem of pattern recognition, for which the technology is evolving rapidly.

A final question concerns the cost-effectiveness of the connectomic approach. Is it to be the exclusive province of a few deep-pocketed laboratories? Here, the answer is clear: the researchers involved have stressed that the connectomic reconstructions will be public resources, which can be used for different purposes by anyone. To be useful, an archive may need to be accompanied by a user-oriented interface, because computer code this complex will be hard for workers other than its creators to use. The effort involved in creating and curating such a public resource should be worth it, because the archives could be the biggest contribution of this work to neuroscience.

Very many structural problems could be attacked using the same original material. ■

Richard H. Masland *is in the Departments of Ophthalmology and Neurobiology, Harvard Medical School, Boston, Massachusetts 02114, USA.*
e-mail: richard_masland@meei.harvard.edu

1. Helmstaedter, M. *et al. Nature* **500,** 168–174 (2013).
2. Takemura, S. *et al. Nature* **500,** 175–181 (2013).
3. Maisak, M. S. *et al. Nature* **500,** 212–216 (2013).
4. Masland, R. H. *Neuron* **76,** 266–280 (2012).
5. Wässle, H., Puller, C., Müller, F. & Haverkamp, S. *J. Neurosci.* **29,** 106–117 (2009).
6. Reichardt, W. in *Sensory Communication* (ed. Rosenblith, W. A.) 303–317 (MIT Press, 1961).
7. Barlow, H. B. & Levick, W. R. *J. Physiol. (Lond.)* **178,** 477–504 (1965).

SOLAR SYSTEM

Saturn's tides control Enceladus' plume

Data obtained by the Cassini spacecraft show that the plume of ice particles at the south pole of Saturn's moon Enceladus is four times brighter when the moon is farthest away from the planet than when it is closest. SEE LETTER P.182

JOHN SPENCER

An analysis of images of Saturn's moon Enceladus taken from the Cassini Saturn orbiter, reported by Hedman *et al.*[1] on page 182 of this issue, shows that the output of the giant plume of ice particles, which jets out of fractures in the south polar region of the moon, is controlled by diurnal changes in the tidal stresses from Saturn*. The authors find a remarkably strong and simple

*This article and the paper under discussion[1] were published online on 31 July 2013.

relationship between the brightness of the plume and Enceladus' position in its orbit around Saturn, providing dramatic confirmation of predictions made in 2007 from a tidal-stress model[2].

Tidal forces from Saturn are the ultimate power source for the extravagant geological activity on Enceladus[3], a small ice world just 500 kilometres in diameter. Enceladus' 1.37-day orbit around Saturn is slightly eccentric, as a result of the periodic gravitational influence of the larger moon Dione. The daily variations in the tidal stresses from Saturn due to
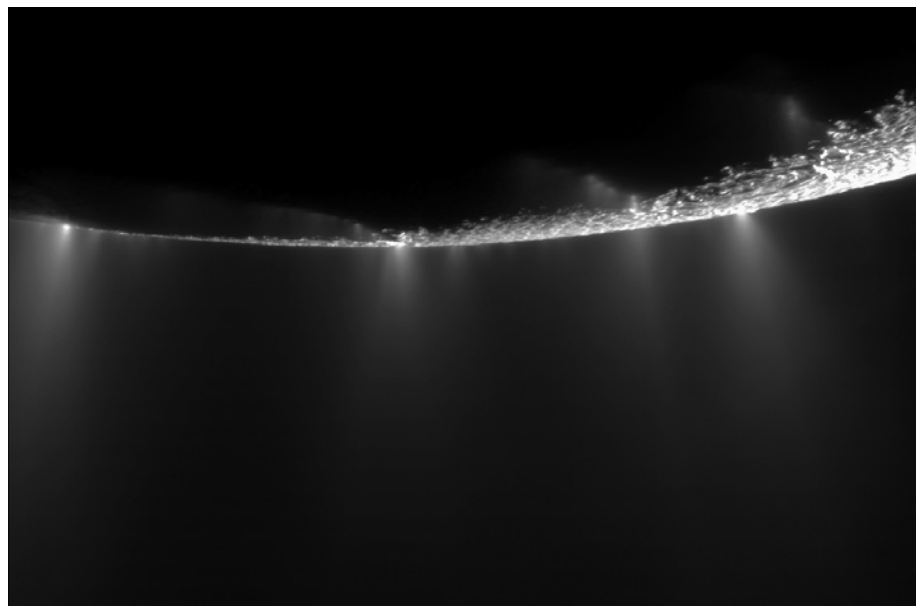
Figure 1 | Enceladus' ice-particle jets. The image shows jets of ice particles emerging from the four 'tiger-stripe' fractures at Enceladus' south pole, photographed in visible light by cameras on the Cassini Saturn orbiter in 2009. Hedman and colleagues' analysis[1] of hundreds of lower-resolution infrared images, taken by Cassini's VIMS instrument, of the plume that results from the combination of these jets, reveals that the plume responds to daily changes in the tidal stresses from Saturn on the fractures, brightening when the fractures are in tension.

SPACE SCIENCE INSTITUTE/JPL/NASA

NATURE

50 Years Ago

A hazard to the animal that has always accompanied war is when persons unaccustomed to animals are drafted to animal units; in the Second World War this was a very common occurrence, because by then, so few persons had experience with horses. Many loved the job and did it well, but others were probably disappointed in a non-mechanical role and not very efficient. Mules reported to be vicious were often underworked and overfed: in any event for a tyro to be placed in charge of an army mule could be described as an experience.
From *Nature* 10 August 1963

100 Years Ago

A very remarkable red-water phenomenon is at present observable in a small pond in Broad Oak Park, Worsley, near Manchester, just in front of the seventh tee on the golf course. The surface of the pond — at any rate at times — is covered in places with an almost blood-red scum, which seems to float on the surface film like fine dust. The scum sometimes assumes a greenish hue. Microscopical examination shows that it is due to the presence of immense numbers of a large species of Euglena, the green chlorophyll of which … is more or less replaced by red haematochrome … Since writing the above I have been able to observe how the Euglenae reach the surface of the water. They evidently secrete some sort of slime in which they become entangled. Bubbles of oxygen gas, given off by the Euglenae in the presence of sunlight, are also caught in this slime, and when these reach a certain size they rise to the surface, trailing strings of slime, with numerous entangled Euglenae, after them.
From *Nature* 7 August 1913

---

this eccentricity distort Enceladus and dump gigawatts of frictional heat into its interior. This heat powers the ejection of the plume of water vapour and ice particles, discovered by Cassini in 2005, from four parallel warm fractures called 'tiger stripes', which surround the moon's south pole (Fig. 1).

Enceladus is thus one of the few places beyond Earth where we can watch geology happen in real time, giving us a primer for understanding other, less active, icy worlds. The ice particles in the plume are salty, hinting at a subsurface source of liquid water and a probable subsurface ocean[4], and the plume gases include complex hydrocarbons and other organic compounds[5]. The likely presence of liquid water and complex organic chemistry makes Enceladus especially intriguing as a potential habitat for extraterrestrial life, providing additional motivation for investigating its interior.

The 2007 modelling study[2] noted that the rhythmic tidal stresses that power Enceladus' activity would also put the tiger-stripe fractures into alternating states of tension and compression during each orbit of Saturn. The study predicted that the majority of the tiger stripes would be in compression when Enceladus is closest to Saturn (periapse), and in tension when it is farthest away (apoapse). Tensional stresses would plausibly open pathways for the venting of plume gases and particles, and thus increase plume activity at apoapse.

Cassini has so far made 20 close fly-bys of Enceladus to investigate its surface and interior and to sample its plume. But the plume is large enough and bright enough to be seen by remote sensing from longer range, permitting more frequent study both with Cassini's visible-wavelength cameras and with its Visual and Infrared Mapping Spectrometer (VIMS). Hints of the predicted variability in individual plume jets were seen in an early study[6], using a relatively small number of images from Cassini's cameras. Hedman and colleagues' work, which is based on VIMS infrared images, was able to reach more definitive conclusions, thanks to systematic analysis of the spectrometer's much larger data set of 252 plume images.

The large data set allowed the team to tease apart orbit-related temporal variations in plume brightness from possible brightness variations caused by long-term changes and by illumination geometry (the micrometre-sized ice particles in the plume brighten greatly when backlit). The temporal variations revealed by this analysis are simple and dramatic — the plume is consistently about four times brighter when Enceladus is at apoapse than when it is at periapse, just as predicted by the 2007 model[2]. This provides strong evidence that the tiger-stripe fractures do indeed open and close each day in response to diurnal tidal stresses, controlling the plume in the process.

Geology, dealing as it does with the complex behaviours and long memories of materials in the solid state, tends to be a messy business. So it is always startling and instructive when simple patterns like this emerge. The strength of the plume–orbit relationship is a strong constraint on how Enceladus works, and a valuable guide to future research. For instance, the 2007 tidal-stress model treats Enceladus' crust as a thin elastic shell that is detached from the interior by a fluid layer, such as a global ocean. Global-ocean models have fallen out of favour for Enceladus, because it is difficult to keep a global ocean from freezing[7], and a regional south polar ocean[8] is now considered more likely. Such refined models can now be tested and constrained by their ability to match the observed plume behaviour.

Also, Cassini's optical cameras, with 50 times higher spatial resolution than that of VIMS, can provide more precise constraints on models of internal structure by tracking the variability of the individual jets that make up the plume[9]. Jet variability can thus provide knowledge of local stresses on the fractures from which the jets emanate, giving another means of investigating local subsurface conditions. The VIMS study itself hints at changes in plume ejection speed at different positions in the orbit, offering another handle on how the plume reaches the surface. No other world beyond Earth allows such detailed analysis of active geophysical processes.

Enceladus displays other strikingly clear-cut patterns. For example, its several geological provinces, defined by age and style of surface deformation, are arranged with almost perfect symmetry around its spin axis and the direction of Saturn[10]. Equally strange is the geometric simplicity of the four active tiger-stripe fractures, all roughly the same length (about 130 km), and evenly spaced about 35 km apart. As with the plume behaviour, such simple patterns must point to important truths, but these other puzzles remain mysterious, and definitive explanations await future research. ■

John Spencer *is at Southwest Research Institute, Boulder, Colorado 80302, USA.*
e-mail: spencer@boulder.swri.edu

1. Hedman, M. M. *et al. Nature* **500,** 182–184 (2013).
2. Hurford, T. A., Helfenstein, P., Hoppa, G. V., Greenberg, R. & Bills, B. G. *Nature* **447,** 292–294 (2007).
3. Spencer, J. R. & Nimmo, F. *Annu. Rev. Earth Planet. Sci.* **41,** 693–717 (2013).
4. Postberg, F., Schmidt, J., Hillier, J., Kempf, S. & Srama, R. *Nature* **474,** 620–622 (2011).
5. Waite, J. H. *et al. Nature* **460,** 487–490 (2009).
6. Spitale, J. N. & Porco, C. C. *Nature* **449,** 695–697 (2007).
7. Roberts, J. H. & Nimmo, F. *Icarus* **194,** 675–689 (2008).
8. Collins, G. C. & Goodman, J. C. *Icarus* **189,** 72–82 (2007).
9. Porco, C., DiNino, D. & Nimmo, F. *44th Lunar Planet. Sci. Conf.* abstr. 1775 (2013).
10. Crow-Willard, E. & Pappalardo, R. T. *EPSC Abstr.* **6,** EPSC-DPS2011-635-1 (2011).

BIOCHEMISTRY

Curbing the excesses of low demand

Metabolic processes are regulated by the relative need for the end product, but this control mechanism may fail if demand is very low. A safety mechanism that copes with low demand has been discovered in bacteria. SEE LETTER P.237

ATHEL CORNISH-BOWDEN

In bacterial cells, the overproduction of metabolites is normally avoided by mechanisms that are similar in principle to control systems in engineering. In this issue, Reaves et al.[1] (page 237) report what happens in mutant bacteria that lack a supposedly essential control mechanism to prevent excessive production of pyrimidine nucleotides — metabolites that act as building blocks for the synthesis of genes, but which are potentially toxic if allowed to accumulate. Instead of observing the expected accumulation, the authors discovered a mechanism in which excess nucleotides are eliminated. In so doing, they identified a plausible role for an enzyme whose physiological function had hitherto been unknown*.

We all know that a kitchen sink is liable to overflow if the tap is left on with the plughole blocked. In most domestic sinks this danger is averted, at least partially, by having an overflow outlet near the top. But in more elaborate engineered systems, such as the domestic toilet, an overflow is avoided by means of negative feedback: as soon as the water in the tank reaches a certain level, the inflow is switched off. Bacterial metabolism is in many ways similar, in that feedback mechanisms prevent the potentially harmful build-up of metabolites. The great explosion of interest in biological regulatory mechanisms in the 1960s followed the realization that negative feedback in metabolism operates in the same way[2,3] as in engineered systems, by allowing the output of an end product to match demand[4] (Fig. 1a).

From an evolutionary perspective, what matters in metabolism is not so much the 'overflow' that occurs in the absence of biological feedback controls, but the build-up of compounds that arises when the requirement for a metabolite is too low to be handled adequately by feedback inhibition (Fig. 1b). This is difficult to test experimentally, but one can mimic the situation by studying mutant bacteria in which the feedback mechanism is suppressed. For example, mutants of *Escherichia coli* bacteria that are insensitive to feedback inhibition of the production of the amino

acid lysine have been studied[5], and behaved as expected: the cells grew poorly, probably as a result of lysine accumulation. By contrast, Reaves et al. observed that the expected overflow does not occur in mutant *E. coli* lacking feedback inhibition of the first enzyme in the biosynthetic pathway leading to pyrimidine nucleotides. Furthermore, the authors saw little effect of this lack of inhibition on the levels of the pathway's end product, cytidine triphosphate (CTP), and they found that the growth rates of the bacteria were normal, except under energy-limiting conditions.

CTP is essential for the synthesis of genes. Although it is produced from aspartate, an amino acid, it is about as different in structure from aspartate as a metabolite can be. Enzymes are normally inhibited by molecules that resemble their substrates, and the discovery that the first step of CTP biosynthesis — the one that uses aspartate as a substrate — is inhibited by CTP was one of the factors that led to the recognition of feedback regulation in metabolism[2].

Surprisingly, Reaves and co-workers observed that neither CTP nor the biosynthetic intermediates leading to it accumulated in the mutant bacteria that lacked feedback

regulation. So why not, and what happened to the excess CTP produced? The authors found that it was degraded and excreted in the form of uracil (an RNA base). They describe this mode of regulation as "directed overflow metabolism"; others have called it "catabolic demand"[4]. The degradation involved a phosphatase enzyme that is evolutionarily conserved, but whose function was previously unknown.

Metabolic regulation is most easily analysed in economic terms of supply and demand[4,6], especially given that the primary function of feedback inhibition is to regulate metabolite concentrations, rather than fluxes. Biological responses to the demand for metabolic end products are common in many systems, thus explaining the occurrence of cooperative feedback inhibition, such as that by CTP in pyrimidine biosynthesis. But the degradation of CTP to uracil observed by Reaves et al. is not a response to demand for uracil, because there is no particular demand for it. Instead, this degradation is a response to the excessive supply of nucleotides, and allows the concentrations of CTP and its biosynthetic precursors to be kept essentially constant when demand for CTP is low (Fig. 1c). Regulation according to supply occurs in biological detoxication pathways, and also in central metabolism. In particular, the mammalian liver does not phosphorylate glucose to form glycogen — the polymer used for energy storage in mammals — to satisfy its own demand for energy, but to maintain glucose homeostasis in other organs by maintaining a constant concentration of glucose in the blood[7].

The CTP-production pathway is therefore an example of a system in which regulation by demand for end product occurs side-by-side with regulation by degradation and excretion of excess end products. Evolution cannot have generated and conserved the excretion mechanism purely to compensate for the artificial
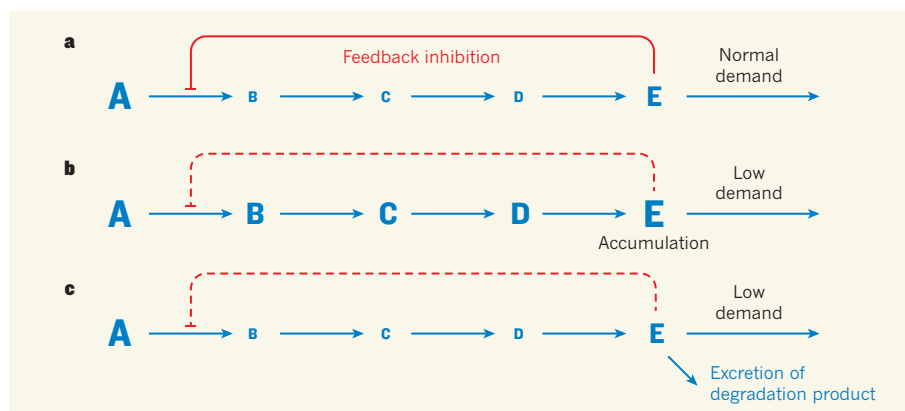
Figure 1 | Regulating metabolite production. a, In this general metabolic pathway, a starting material (A) is converted to an end product (E) via intermediates (B to D). When demand for E is normal, regulation of the pathway occurs by feedback inhibition of the first step — E inhibits the conversion of A to B. The sizes of the letters indicate the amount of each compound that accumulates. **b**, At very low demand, feedback inhibition is ineffective (broken line). Production of E is therefore excessive, and so it and the intermediates B to D accumulate. **c**, Reaves et al.[1] report that homeostasis of end products can be maintained when demand is very low if any excess is degraded and the resulting products are excreted. The authors discovered such a mechanism in the metabolic pathway that produces CTP, a pyrimidine nucleotide.

*This article and the paper under discussion[1] were published online on 31 July 2013.

deletion of feedback in experiments, so this pathway must represent a back-up strategy for physiological states in which demand falls below the finite range within which feedback is effective. If so, then we should expect to find examples of the same sort of behaviour in other pathways.

One other point will have occurred to alert readers. If excess production of pyrimidine nucleotides is overcome by converting the excess into uracil and excreting it, this implies that nutrients are not being used efficiently by the bacteria. That is why the mutant bacteria grew as well as normal ones when energy supplies were abundant, but somewhat more slowly when energy was limited. ∎

Athel Cornish-Bowden *is at Bioénergétique et Ingénierie des Protéines, Institut de Microbiologie de la Méditerranée, CNRS Aix-Marseille Université, BP 71, 13402 Marseille Cedex 20, France. e-mail: acornish@imm.cnrs.fr*

1. Reaves, M. L., Young, B. D., Hosios, A. M., Xu, Y.-F. & Rabinowitz, J. D. *Nature* **500,** 237–241 (2013).
2. Umbarger, H. E. *Science* **123,** 848 (1956).
3. Yates, R. A. & Pardee, A. B. *J. Biol. Chem.* **221,** 757–770 (1956).
4. Hofmeyr, J.-H. S. *Essays Biochem.* **45,** 57–66 (2008).
5. Boy, E. & Patte, J.-C. *J. Bact.* **112,** 84–92 (1972).
6. Hofmeyr, J.-H. S. & Cornish-Bowden, A. *FEBS Lett.* **476,** 47–51 (2000).
7. Cárdenas, M. L. *"Glucokinase": Its Regulation and Role in Liver Metabolism* (Landes Bioscience, 1995).

EVOLUTIONARY BIOLOGY

The handiwork of tinkering

A comparison of regulatory DNA sequences in humans, macaques and mice during embryonic limb development reveals thousands of sites of enhanced regulatory activity that are likely to have driven the evolution of our characteristic anatomy.

PAUL FLICEK

Understanding what makes us human is a long-standing research goal that has been made more accessible in the genomic era. It is clear that specific features of human anatomy are encoded in the genome, and the extent of gene sharing among mammals suggests that most of these features result from gene-regulatory changes that occurred at some point in the human lineage. Writing in *Cell*, Cotney *et al.*[1] have provided the first comprehensive map of the genomic regions responsible for the unique morphology of human limbs. They also show that the vast majority of gains of regulatory activity in the human lineage occur in genomic regions that trace back to the base of the mammalian tree. These results demonstrate once again that, as Nobel-prizewinning biologist François Jacob described it[2], evolution is a tinkerer using the material at hand to create something workable.

Compared with other mammals and primates, humans have several specific limb characteristics, including the relative lengthening of the thumb and multiple adaptations of the foot to support walking upright. These and other developmental details are evident early in embryonic development, and so the relevant regulatory regions of our genomes must be active at specific developmental times to drive the appropriate physiology.

Cotney and colleagues set out to map these regions in the human genome using comparative regulatory genomics — the recently developed technique of performing the same functional experiment in matched tissues of multiple species to gain insight into regulatory evolution[3]. What makes this study so significant is that the experiments were conducted on embryonic limb tissue, with careful matching of multiple developmental stages in samples from humans, macaques and mice. At each stage, the authors identified genomic locations at which the lysine amino-acid residue at position 27 of histone 3 proteins (histones are proteins around which DNA is wound) was modified by acetylation. This mark, referred to as H3K27ac, identifies[4] active gene-regulatory regions, such as promoters and enhancers. H3K27ac markings at regions regulating the expression of *HOX* genes, which have a deeply evolutionarily conserved role in development, provided a natural control, and inspire confidence that the profiled tissues were indeed from the same developmental stage across the three species.

These maps of promoter and enhancer activity show the progression of regulatory innovation with evolution. The limb bud at human embryonic day 33 was the earliest stage profiled by Cotney *et al.*, and was also the stage at which the fewest promoter sequences (2%) showed H3K27ac marks in human samples but not at the related (derived from the same ancestor) sequence in macaque tissues — such comparisons were referred to as human-lineage gains. The latest developmental stage profiled

Figure 1 | The evolutionary origins of mammalian and human limb development. Cotney *et al.*[1] studied limb tissue from mice, macaques and humans at various stages of embryonic development (equivalent to human embryonic day 33 (E33), E41, E44 and E47) to identify active gene-regulatory sequences (enhancers and promoters). **a,** Their results show that, as development proceeds, the fraction of human enhancers that were active in human tissue only (human-gain enhancers) increased, demonstrating the specific needs of human limb development. The other categories are enhancers that were active in all three species (stable mammalian enhancers) and enhancers that are present but not active in all three species (orthologues). **b,** The study also reveals that many of the sequences that are stably active in the three mammals are also present in non-mammalian vertebrate genomes, whereas the human-gain enhancers are, on average, younger.

by the authors, corresponding to human embryonic day 47, focused on the hand and foot plates; at this stage, Cotney *et al.* identified 16% of promoters with human-lineage gains. These results are in accordance with the intuitive idea that later developmental stages will demonstrate more human-specific morphological features (Fig. 1a). In total, more than 2,000 promoters and nearly 3,000 enhancers showed significant human-lineage gains at at least one assayed time point, demonstrating the dramatic orchestration of gene expression that is required for human limb development.

In their comparisons between species, the authors classified enhancer regions into three categories: orthologous (an enhancer sequence that simply exists in all three species); those that exist and are stably marked with a consistent level of H3K27ac in all three species; and those that exist in all three species but that show a gain of H3K27ac in the human lineage.

The characteristics of the sequences that fit into these classes, including their evolutionary age and conservation, provide some seemingly surprising results. First, the majority of human-lineage gains do not occur in highly conserved elements, although they do not seem to be completely unconstrained; rather, the greatest sequence conservation is seen in the stably marked regions. The stably marked and orthologous regions are also significantly older, in evolutionary terms, than those in which activity arises in the human lineage, with many also being present in non-mammalian vertebrates (Fig. 1b). Human-lineage gains, on the other hand, tend to be found in sequences that arose between the time that marsupials and placental mammals shared a common ancestor and the divergence of the placental lineage. A few of the sequences identified as human-lineage gains also overlap with regions previously identified as showing significant change since the divergence of the human lineage from chimpanzees[5]. Cotney and colleagues' findings provide clues to the roles of 16 of these 'human accelerated regions', and these may now be attractive candidates for further analysis.

Exactly what is going on in the regions with enhanced regulatory activity in the human lineage is still not known. The sites had no obvious enrichments for specific transcription-factor-binding motifs. Also absent were specific repetitive elements, which have been shown to contribute to the regulatory rewiring of multiple mammalian lineages[6] and to be a contributing factor to the evolution of pregnancy[7]. Understanding the molecular drivers at these and other certain-to-be-discovered regions of human regulatory change is both a formidable and an exciting challenge.

There are limitations to Cotney and colleagues' analysis. For example, the quality of the genome sequences varies between the species, which may have contributed to the authors' inability to find orthologous regions in macaques for some human sites. Alternative explanations

for this lack of orthologues, based on changes in copy number or human-lineage duplications[8], are also complicated by the absence or possible misassembly of these regions in the human genome sequence. Moreover, there are problems associated with mapping and analysis of the short DNA sequences resulting from the ChIP-seq analysis used by the authors to sequence regions containing H3K27ac marks. However, the study provides clear insight into the regulatory changes that help to make us human, and the authors have presented an extremely valuable map, connecting regulatory regions and gene-expression changes involved in human limb development. ∎

Paul Flicek *is in the European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK.*
e-mail: flicek@ebi.ac.uk

1. Cotney, J. *et al. Cell* **154,** 185–196 (2013).
2. Jacob, F. *Science* **196,** 1161–1166 (1977).
3. Schmidt, D. *et al. Science* **328,** 1036–1040 (2010).
4. Creyghton, M. P. *et al. Proc. Natl Acad. Sci. USA* **107,** 21931–21936 (2010).
5. Lindblad-Toh, K. *et al. Nature* **478,** 476–482 (2011).
6. Schmidt, D. *et al. Cell* **148,** 335–348 (2012).
7. Lynch, V. J., Leclerc, R. D., May, G. & Wagner, G. P. *Nature Genet.* **43,** 1154–1159 (2011).
8. Marques-Bonet, T., Girirajan, S. & Eichler, E. E. *Trends Genet.* **25,** 443–454 (2009).

CLIMATE SCIENCE

Solution proposed for ice–age mystery

The ice sheets retreated 10,000 years ago during a peak in solar radiation, but this peak was no larger than previous ones. A modelling study suggests why the ice sheets were unusually vulnerable to melting at that time. SEE LETTER P.190

SHAWN J. MARSHALL

I first encountered *Ice Ages: Solving the Mystery*[1], the seminal book by John and Katherine Imbrie, as an undergraduate student, and it played no small part in drawing me in to graduate studies on ice-age climate dynamics. Imbrie *père et fille* describe the various strands of evidence establishing that Earth–Sun orbital variations are the main driver of glacial cycles: the recurring flow and ebb of ice sheets over the continents during an ice age. About 40 such glacial cycles have shaped our planet over the past 2.6 million years (the Quaternary period), representing the most dramatic example of climate variability in Earth's recent history.

But there is one nagging problem: as much as Earth's orbital wobbles seem to pace the advance and retreat of ice sheets, many aspects of ice-age climate dynamics remain a mystery. For one thing, those who model climate and ice sheets have not yet been able to simulate glacial cycles in a realistic way. Glacier advance into mid-latitudes requires severe cooling and increased snowfall compared with present-day conditions, to an extent that far exceeds the predicted response of the Earth system to 'cold' orbital configurations in climate models. It is even more difficult to get rid of continental ice sheets once they gain a foothold on the landscape. The modelling results reported by Abe-Ouchi *et al.*[2] in this issue may provide a solution to these problems.

The crux of the challenge in modelling glacial cycles is that Earth's response to orbital

forcing is entirely out of proportion. Changes in Earth's tilt axis and the eccentricity of its orbit around the Sun give rise to geographical and seasonal changes in incoming solar radiation. The global annual impact of these variations is negligible, but what really matters to the ice sheets is the amount of sunlight at high northern latitudes during the summer melt season. Peak radiation in this region varies by up to 100 watts per square metre because of orbital variations (Fig. 1a); this would certainly affect Arctic ice cover. However, integrated summer radiation, which is what counts in ice-sheet melting[3], has deviated by less than 10% from present-day values over the most recent glacial cycle (Fig. 1b), and it is not obvious why this has elicited such a large shift in global climate.

In fact, a host of positive feedbacks — cooling influences associated with increases in snow and ice cover — conspire to amplify the orbital signal and send the world careering into glaciations. It is difficult to overcome these cooling influences, and so orbital changes alone are not enough to trigger deglaciation. The most recent glaciation persisted for roughly 100,000 years, and the ice sheets survived several periods of orbital warming before they finally destabilized and withdrew, starting about 20,000 years ago (Fig. 1c). At that time, summer solar radiation in the Northern Hemisphere increased, eventually peaking at about 6% above modern levels 10,000 years ago. But similar peaks occurred earlier during this period of glaciation, so what was different about this one?
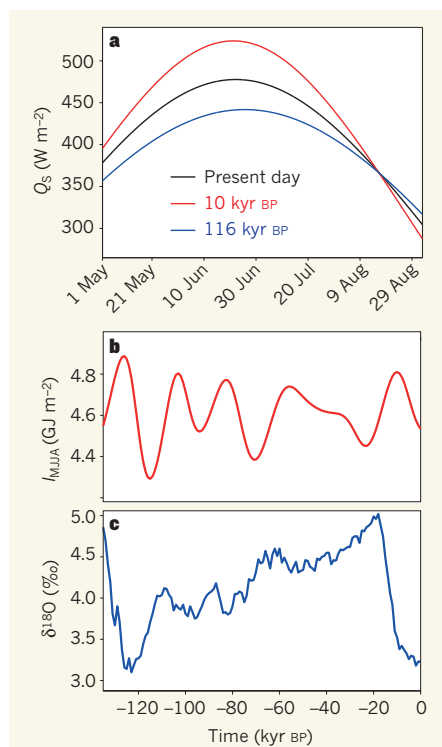
Figure 1 | Solar radiation and ice-sheet coverage.
a, The average daily incoming solar radiation (Q_s) at 60° N from May to September varies as a result of fluctuations in Earth's orbit around the Sun, as revealed by these data for 116,000 years before present (116 kyr BP; the inception of the last glacial period), 10 kyr BP (the period of maximum insolation during the most recent deglaciation) and the present day[3]. **b**, The integrated summer insolation (I_{MJJA}) at 60° N during the last glacial cycle reveals several peaks. **c**, Stacked benthic stable isotope ratios ($\delta^{18}O$) from the global ocean are a proxy for global ice-sheet volume during the glacial–interglacial cycle[7]. Comparison of **b** with **c** reveals that the insolation peak that triggered deglaciation was only as large as other insolation peaks that did not induce deglaciation. Abe-Ouchi et al.[2] report that the geometry of North America and the time taken for bedrock to sink beneath ice sheets explain why deglaciation occurred when it did.

Through asynchronous coupling of sophisticated climate and ice-sheet models, Abe-Ouchi and co-authors make a convincing case that the geometry of North America and the long response time of isostatic compensation — the change in height of Earth's surface in response to ice-sheet formation and retreat — are the main agents that transform 19,000-year (19-kyr), 23-kyr and 41-kyr orbital variations into a 100-kyr Earth-system response[4]. Ice sheets build up and flow southwards in both North America and Eurasia, taking many millennia to thicken and advance to their southern limits. Subglacial bedrock is depressed as underlying mantle material flows slowly outwards. At equilibrium, a 3,000-metre-thick ice sheet undergoes about 1,000 metres of subsidence[5], but achieving equilibrium takes thousands of years. Similarly, land that was underneath the glacial ice

sheets is still springing back.

Isostatic subsidence is one of the few negative feedbacks associated with glaciation: as an ice sheet slowly sinks, its surface lowers into a warmer climate, increasing the amount of melt and the area of the ice sheet exposed to melting. In Abe-Ouchi and colleagues' simulations, this process becomes most effective late in the glacial cycle, when the North American ice sheets are thick and have advanced far enough south; because this takes a long time, North America is set for a 100-kyr response. By contrast, the geography of the Eurasian ice sheets (which are thinner and less extensive, and occur in a warmer climate) gives them less inertia, and so they are more sensitive to 20- and 41-kyr orbital variations.

This idea is not new — earlier modelling studies[5,6] also implicated isostatic rebound as one of the main processes underlying the 100-kyr glacial cycle. However, free-running simulations of the cycle have never before been achieved without invoking 'exotic mechanisms' — such as imposed ocean-circulation changes, dynamic ice-sheet destabilization or 'dusting' of the ice sheets — that force deglaciation at the 'right' time. One innovative technique that helps to capture the glacial cycle in Abe-Ouchi and colleagues' analysis is the use of multiple snapshots from climate models, which provide information about different ice-sheet sizes, carbon dioxide concentrations and orbital configurations. This is necessary because the computational time required to run a sophisticated climate model over tens of millennia is still prohibitively long.

However, some lingering mysteries remain,

such as the effects of the oversimplified treatment (or absence) of ice sheet–ocean interactions, basal flow (ice-sheet sliding and subglacial sediment deformation) and ice-stream processes in the authors' simulations. Furthermore, ice-sheet melt rates are estimated only from air temperature, and are not based on energy-balance physics within the atmospheric model used by the authors. As climate and ice-sheet models become more sophisticated, we will see further refinement of these results.

Moreover, Abe-Ouchi and colleagues' findings do not explain the transition that took place 900,000 years ago, when the world moved from 41-kyr to 100-kyr glacial cycles. Isostatic time scales and North American geography did not change across this boundary, so another factor must have been at work. There are some layers yet to be explored in the mysteries of the ice age. ■

Shawn J. Marshall *is in the Department of Geography, University of Calgary, Calgary, Alberta T2N 1N4, Canada, and at the Canadian Institute for Advanced Research.*
e-mail: shawn.marshall@ucalgary.ca

1. Imbrie, J. & Imbrie, K. P. *Ice Ages: Solving the Mystery* (Harvard Univ. Press, 1986).
2. Abe-Ouchi, A. *et al. Nature* **500**, 190–193 (2013).
3. Huybers, P. *Science* **313**, 508–511 (2006).
4. Berger, A., Mélice, J. L. & Loutre, M. F. *Paleoceanography* **20**, PA4019 (2005).
5. Peltier, W. R. & Hyde, W. T. in *The Physical Basis of Ice Sheet Modeling* 247–260 (Int. Assoc. Hydrol. Sci., 1987).
6. Oerlmans, J. *Nature* **287**, 430–432 (1980).
7. Lisiecki, L. E. & Raymo, M. E. *Paleoceanography* **20**, PA1003 (2005).

PALAEONTOLOGY

Jurassic fossils and mammalian antiquity

Two new Jurassic fossils yield conflicting reconstructions of the mammalian tree. These divergent genealogies have profoundly different implications for the origin and early diversification of mammals. SEE ARTICLE P.163 & LETTER P.199

RICHARD L. CIFELLI & BRIAN M. DAVIS

Fossil discoveries and molecular studies in recent decades have greatly advanced our understanding of mammalian relationships and diversification[1]. Yet major points of disagreement remain over some of the basal branches of the family tree. There is little doubt that mammals, strictly defined[2], were widespread and ecologically diverse by the middle of the Jurassic period, about 165 million years ago[3–5]. But when did they originate? A major sticking point is the inclusion (or not)

of certain poorly known early forms. Substantial information is now provided by two separate discoveries, reported in this issue by Zhou *et al.*[6] (page 163) and Zheng *et al.*[7] (page 199), of splendidly preserved fossils from China that date to between 160 million and 165 million years ago.

Both fossils, which include evidence of fur but lack complete skulls, have been assigned to the Haramiyida. This enigmatic group includes fossils dating back to the Late Triassic — that is, about 40 million to 50 million years before the appearance of undoubted

mammals[4,5]. Intriguingly, haramiyidans have some rodent-like specializations of the teeth, and in this respect they vaguely resemble members of another extinct group, the Multi-tuberculata, which is well known and has consistently been placed firmly within the mammalian tree[2,3]. Although the new fossils seem to be related, detailed examination reveals significant differences between the two animals — differences that lead the authors to contrasting views of mammalian relationships which imply quite different models for the origin and initial diversification of these iconic organisms.

Zheng and colleagues describe *Arboroharamiya*, a short-faced presumed omnivore or herbivore that was adapted for life in the trees (as implied by the name). Notably, the authors interpret the back of the jaw to be simple (although this is not clear from the photographs in the paper's Supplementary Information), indicating that the bones of the middle ear had assumed a typical mammalian configuration[8]. Surprisingly, the results of their analysis place haramiyidans (including *Arboroharamiya*) and multituberculates together within Mammalia (Fig. 1a). This implies that mammals originated at least 215 million years ago — a much earlier date than many palaeontologists would accept[2,3], but one that is in agreement with a recent estimate based on a molecular-clock model[9].

Megaconus, described by Zhou and colleagues, comes from the same geological formation as *Arboroharamiya*, but is slightly older. Apparently adapted for terrestrial life and a plant-based diet, *Megaconus* is primitive in many respects: the counterparts of the mammalian middle-ear bones remain attached to the jaw, for example, and the ankle resembles that of pre-mammalian forms. The authors' analysis places haramiyidans (including *Megaconus*) outside Mammalia and unrelated to multituberculates (Fig. 1b). Under this interpretation, the minimal age for mammalian origin is much younger, and is constrained by the appearance of multiple, uncontested types of mammal by about the Middle Jurassic[3,4].

The evolutionary history of various mammalian groups has been characterized as 'long fuse', 'short fuse' and 'explosive'[1,9,10], depending on the time elapsed between their origin and significant diversification. Zheng and colleagues' tree, which places the origin of mammals at 40 million to 50 million years before their Jurassic radiation, implies a long-fuse model for the group as a whole (Fig. 1a). Zhou and colleagues' competing hypothesis, on the other hand, is consistent with an explosive model (Fig. 1b). These alternative family trees and implied models of diversification have profound ramifications for interpreting the significance of key adaptations, ecological interactions and other events in the early history of mammals[3,8,11].

Neither family tree perfectly explains all the
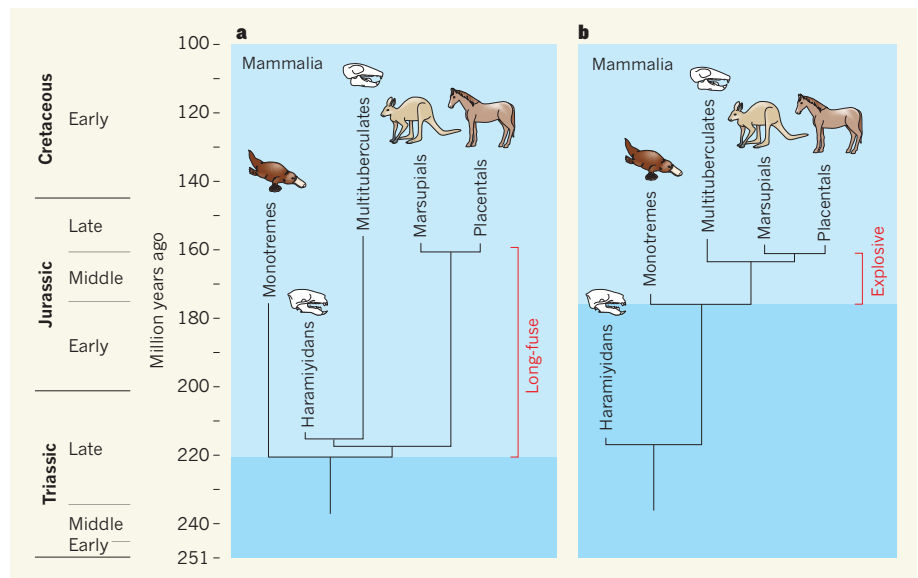
Figure 1 | Alternative interpretations of early mammalian history. Discoveries identified as uncontested mammals appear in the fossil record by the beginning of the Middle Jurassic (175 million years ago). Close relatives have been found in rocks dating to the Late Triassic (about 215 million years ago), but most palaeontologists place these groups outside Mammalia[2]. Zhou et al.[6] and Zheng et al.[7] describe fossils that are both assigned to one such group, the Haramiyida. Although the fossils are relatives, the two teams reach different conclusions as to where haramiyidans fit in the scheme of mammalian relationships, with strikingly different implications for the origin and diversification of mammals. **a**, Zheng et al. nest haramiyidans within Mammalia, implying a Late Triassic origin for mammals and a long-fuse model[10] of mammal diversification. (Early mammalian splits in **a** are extrapolated from the first appearance of haramiyidans[4] and a marsupial–placental split on the basis of the age of the extinct genus *Juramaia*[12]). **b**, By contrast, Zhou et al. place haramiyidans outside Mammalia, suggesting that the initial radiation of mammals occurred explosively during the Middle Jurassic, immediately following their origin. (Sequential mammalian splits in **b** are based on the ages of the extinct genus *Henosferus*[5], the earliest multituberculates[4] and *Juramaia*[12], respectively). Skull reconstructions from refs 4 and 6.

data, of course. Indeed, each implies seemingly unlikely examples of independent evolution or reversal (homoplasy) of key characteristics. Zheng and colleagues' model, for instance, suggests that the complex, three-boned middle ear, long regarded as a diagnostic feature of mammals[2], may have arisen independently at least three times — in the multituberculate–haramiyidan group, in monotremes (platypuses and echidnas) and in therians (marsupials and placentals). Similarly, the tree proposed by Zhou et al. requires independent evolution of strikingly similar tooth features — the presence of a single pair of enlarged, forward-facing incisors and complex cheek teeth with multiple rows of cusps, for example, that are presumably associated with herbivory — in haramiyidans and multituberculates.

These contrasting results, and their implications for mammalian evolution, should be considered in light of the underlying data. Of the two fossils, *Megaconus* is the more complete, with 41% of anatomical characters scored, compared with 23% for *Arboroharamiya*. In addition, the analysis by Zhou et al. is more comprehensive in its sampling of species and anatomy, particularly with respect to haramiyidans and multituberculates.

The obvious next step, of course, is to conduct analyses to synthesize and reconcile the data presented in these two contributions.

Ultimately, however, more and better fossils, ideally including skulls, will be needed to refine knowledge of early mammalian radiations, both in terms of relationships and palaeobiology. ∎

Richard L. Cifelli *is at the Oklahoma Museum of Natural History, Norman, Oklahoma 73072, USA.* **Brian M. Davis** *is in the Department of Anatomical Sciences and Neurobiology, University of Louisville, Louisville, Kentucky 40202, USA. e-mails: rlc@ou.edu; bmdavis81@gmail.com*

1. O'Leary, M. A. et al. Science **339**, 662–667 (2013).
2. Rowe, T. J. Vert. Paleontol. **8**, 241–264 (1988).
3. Luo, Z.-X. Nature **450**, 1011–1019 (2007).
4. Kielan-Jaworowska, Z., Cifelli, R. L. & Luo, Z.-X. Mammals From the Age of Dinosaurs: Origins, Evolution, and Structure (Columbia Univ. Press, 2004).
5. Gaetano, L. & Rougier, G. W. J. Mamm. Evol. **19**, 235–248 (2012).
6. Zhou, C.-F., Wu, S., Martin, T. & Luo, Z.-X. Nature **500**, 163–167 (2013).
7. Zheng, X., Bi, S., Wang, X. & Meng, J. Nature **500**, 199–202 (2013).
8. Luo, Z.-X. Annu. Rev. Ecol. Evol. Syst. **42**, 355–380 (2011).
9. Meredith, R. W. et al. Science **334**, 521–524 (2011).
10. Archibald, J. D. & Deutschman, D. H. J. Mamm. Evol. **8**, 107–124 (2001).
11. Rowe, T. B., Macrini, T. E. & Luo, Z.-X. Science **332**, 955–957 (2011).
12. Luo, Z.-X., Yuan, C.-X., Meng, Q.-J. & Ji, Q. Nature **476**, 442–445 (2011).

ARTICLE

A Jurassic mammaliaform and the earliest mammalian evolutionary adaptations

Chang-Fu Zhou[1], Shaoyuan Wu[1]†, Thomas Martin[2] & Zhe-Xi Luo[2,3]

The earliest evolution of mammals and origins of mammalian features can be traced to the mammaliaforms of the Triassic and Jurassic periods that are extinct relatives to living mammals. Here we describe a new fossil from the Middle Jurassic that has a mandibular middle ear, a gradational transition of thoracolumbar vertebrae and primitive ankle features, but highly derived molars with a high crown and multiple roots that are partially fused. The upper molars have longitudinal cusp rows that occlude alternately with those of the lower molars. This specialization for masticating plants indicates that herbivory evolved among mammaliaforms, before the rise of crown mammals. The new species shares the distinctive dental features of the eleutherodontid clade, previously represented only by isolated teeth despite its extensive geographic distribution during the Jurassic. This eleutherodontid was terrestrial and had ambulatory gaits, analogous to extant terrestrial mammals such as armadillos or rock hyrax. Its fur corroborates that mammalian integument had originated well before the common ancestor of living mammals.

Clade Mammaliaformes[1]
Order Haramiyida
Family Eleutherodontidae[2,3]
***Megaconus mammaliaformis* gen. et sp. nov.**

Etymology. *Mega* (Greek): large; *conus* (Latin): cusp, after the single hypertrophied anterior cusp of lower premolars; *mammaliaformis*: in reference to ancestral mammalian features.

Holotype. Paleontological Museum of Liaoning at Shenyang Normal University (PMOL) AM00007A and AM00007B (Figs 1, 2 and Supplementary Figs 1–3), preserved with *in situ* dentition, mandibles and associated mandibular middle ear, and most of the postcranial skeleton. Life Science Identifier (LSID): urn:lsid:zoobank.org:act:852BDC75-5D40-4486-990F-2F3682B40FF6.

Locality and age. The Daohugou site in the Tiaojishan Formation of Inner Mongolia, China; the locality was directly dated to be 165–164 million years (Myr) old[4–6]. The mammaliaform assemblage from Daohugou includes the docodont *Castorocauda*[7], the triconodontid *Volaticotherium*[8] and a pseudo-tribosphenic mammal[9]. The Tiaojishan Formation also yielded a eutherian[10] from another locality, at a higher stratigraphic level, which was separately dated to be 161 ± 1.44 Myr[5].

Diagnosis. $I^2\text{-}C^0\text{-}P^2\text{-}M^3/I_1\text{-}C_0\text{-}P_2\text{-}M_3$ (Fig. 2), with a procumbent, enlarged lower incisor I_1 and rhomboidal upper incisor I^2; each lower premolar has an enlarged anterior cusp, and a talonid heel consisting of two cusp rows: P_1 anterior cusp is hypertrophied and re-curved; P_2 talonid is basined. Molars are hypsodont ('high-crowned') by roots[11], and each has multiple roots that are fused proximally but divided distally. Lower molars have two longitudinal rows, each of multiple cusps mostly of pyramidal shape, plus a labial cingulid. Upper molar M^1 has three rows, each of four cusps; M^2 has full labial and middle rows, plus a short lingual row of only two cusps and restricted anteriorly, forming the anterolingual wing. M^3 has only two rows, of which the lingual row is longer. The lingual cusp row on M^1–M^2 is lingually offset from the M^3 (Fig. 2 and Supplementary Fig. 2). Both upper and lower molars are interlocked either by an embayment on the cingulum to receive the preceding tooth, or by tongue-in-groove

interlock on the hypsodont roots. *Megaconus mammaliaformis* is identical to *Eleutherodon*[2] and *Sineleutherus*[12,13] in terms of highly distinctive hypsodont and proximally fused multiple roots, and interlocking features between adjacent molars (Fig. 2 and Supplementary Figs 4, 5). *Megaconus* is different from other eleutherodontids in lacking the dense flutings on the tooth crown, and is distinguishable from all mammaliaforms and cynodonts of the Jurassic and Cretaceous periods by a combination of autapomorphic and plesiomorphic skull and postcranial features, and from all clades of crown mammals by many characteristics (Figs 1 and 3; differential diagnosis in Supplementary Information).

Phylogenetic relationship

Placement of haramiyidans (including *Megaconus* and other eleutherodontids) is broadly relevant to the earliest evolution of mammalian features and timing of origin of basal clades of mammals. The complete dentition, mandible and partial skeleton of *Megaconus* here provide new characteristics for testing alternative phylogenetic hypotheses of haramiyidans, previously based on teeth and their occlusal features[2,3,14–20]. According to our analysis of a matrix of 475 characters and 110 clades (considerably expanded from refs 10, 18), including all observed features in *Megaconus*, and expanded sampling of relevant eleutherodontids, haramiyidans and multituberculates, *Megaconus* is nested within haramiyidans among mammaliaforms, outside of crown Mammalia, and is separated from multituberculates nested in the crown Mammalia (Fig. 4 and Supplementary Fig. 9). The hypothesis that haramiyidans (including *Megaconus*) are a mammaliaform clade outside of the crown Mammalia is the most parsimonious given our data set (Supplementary Information). By nonparametric test, the best strict consensus tree (Supplementary Fig. 9) is 16 steps shorter than, and significantly different ($P \leq 0.0389$) from, the suboptimal 'allotherian' hypothesis that would group haramiyidans with multituberculates in one clade[16,17,20].

This placement of *Megaconus* has resulted primarily from mandibular and postcranial features. *Megaconus* is more plesiomorphic than living

[1]Paleontological Museum of Liaoning, Shenyang Normal University, Shenyang, Liaoning 110034, China. [2]Steinmann-Institut für Geologie, Mineralogie und Paläontologie, Universität Bonn, Nussallee 8, 53115 Bonn, Germany. [3]Department of Organismal Biology and Anatomy, the University of Chicago, Chicago, Illinois 60637, USA. †Present address: Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, Tianjin Medical University, Tianjin 300070, China.
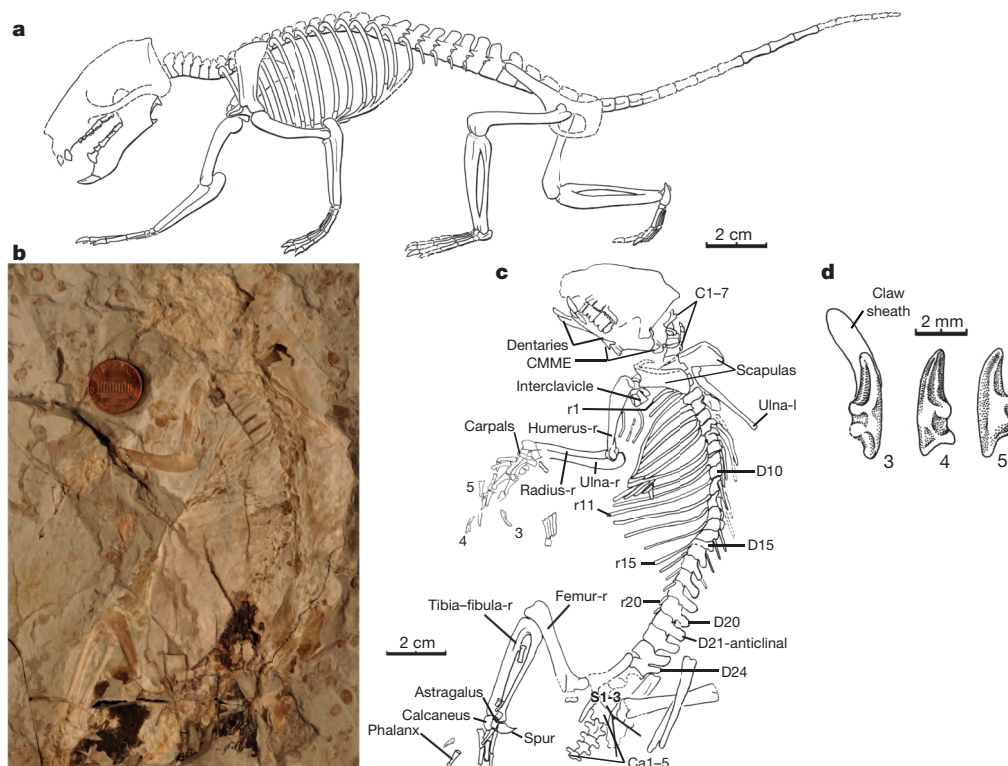
Figure 1 | New Jurassic mammaliaform *Megaconus mammaliaformis*.
a, Skeletal reconstruction. **b**, Holotype counterpart (Paleontological Museum of Liaoning (PMOL)-AM00007B). **c**, Skeletal feature identification; the left (-l) versus right (-r) sides are designated according to the main-part PMOL-AM00007A (Supplementary Fig. 1). **d**, Manual terminal phalanges. Details on dental and skeletal structures can be found in Supplementary Figures. C, cervicals; Ca, caudal vertebrae; CMME, preserved elements of cynodont mandibular middle ear[24]; D, dorsal vertebrae (D1–15 designated as 'thoracic'; D16–24 as 'lumbar'); r, ribs; S, sacral vertebrae; 3–5, the preserved manual terminal phalanges 3–5.

mammals in having the postdentary trough to hold the mandibular middle ear (Fig. 2 and Supplementary Figs 2 and 3), but similar to most mammaliaforms ranging from *Sinoconodon* to docodontans[18,21–26], and to *Haramiyavia*, in having the postdentary trough[15]. It differs from multituberculates in that the latter group has the middle ear fully separated from the mandible[25]. The mandible of *Megaconus* has a distinctive angular process in contrast to the rounded angular region of multituberculates. At the ankle, the calcaneus is positioned side by side with the astragalus, and lacks the long calcaneal tuber and laterally compressed calcaneal body, in contrast to multituberculates that have partial superposition of the astragalus over the compressed calcaneus with an elongate calcaneal tuber (Fig. 3). The astragalotibial joint of *Megaconus* lacks all of the derived features of multituberculates, such as separate medial versus lateral tibial facets to articulate the medial versus lateral distal tibial condyles for a highly mobile and flexible ankle of multituberculates[27].

Implications for dental evolution of early mammals

We note that eleutherodontids have interlocking features between adjacent molars (Supplementary Figs 4 and 5), as in many other (although not all) Triassic–Jurassic mammaliaforms and many crown mammals. Mammals are characterized by only one generation of tooth replacement of antemolar teeth, and their molars usually have fully divided roots and precise upper-to-lower crown occlusion, in contrast to the teeth of most pre-mammalian cynodonts (except for traversodontids and tritylodontids[28]), which have multiple replacements, lack full root division and lack precise occlusion. Fewer replacement(s) at a locus and slower ontogeny of individual teeth are crucial for mammalian teeth to achieve precise upper-to-lower occlusion, the basis for feeding specialization by distinctive and diverse molar crowns[3], and by different growth patterns[11]. These are fundamental for evolutionary diversification of major clades. *Megaconus*

adds to a growing body of evidence that mammaliaforms developed interlocking mechanisms between adjacent teeth for precise alignment of teeth in a tooth row, most probably correlated with the mammal-like and slower ontogeny of a tooth and slower replacement than those of most cynodonts[3].

Teeth of living mammals can develop high enamel crowns, or a high dentine wall above the divided roots, as a consequence of varying growth rates and for specialized feeding adaptation[11]. Most mammaliaforms, such as the haramiyid *Thomasia*, have brachydont (low-crowned) molars with fully divided roots[14]. By contrast, eleutherodontids (Fig. 2, Supplementary Figs 3–5 and Supplementary Video 1) have hypsodont and proximally fused multiple roots for molars, also known as dentine or root hypsodonty, convergent to the dentine hypsodont teeth of some eutherians, among diverse types of hypsodont teeth of living therians[11]. Teeth erupt at a slower rate for a longer duration in hypsodont teeth in general, and the root growth phase of tooth ontogeny is longer for dentine hypsodont molars in particular[11]. Thus dentine hypsodont molars of *Megaconus* indicate that mammaliaforms have a wide range of tooth growth patterns, some of which are convergent to phylogenetically unrelated taxa within living mammals.

Several cynodonts and many mammals have molariform teeth with (usually three) longitudinal, multicusped rows of the upper teeth alternately occluding with (usually two) multicusped rows of the lowers (Fig. 4), with backward power stroke of the mandible for specialized feeding on plants[16,28–30]. Similar (although not identical) dental patterns evolved in traversodontids and tritylodontids among pre-mammalian cynodonts[28] (Fig. 4), in haramiyidans[14–17], in multituberculates[16,29], and prominently in rodent placental mammals[30].

Ever since the discovery of the haramiyidans of the Late Triassic, it has been debated whether haramiyidans and multituberculates belong to a clade[2,3,16,17,31] or whether they are separate lineages[1,14,15].
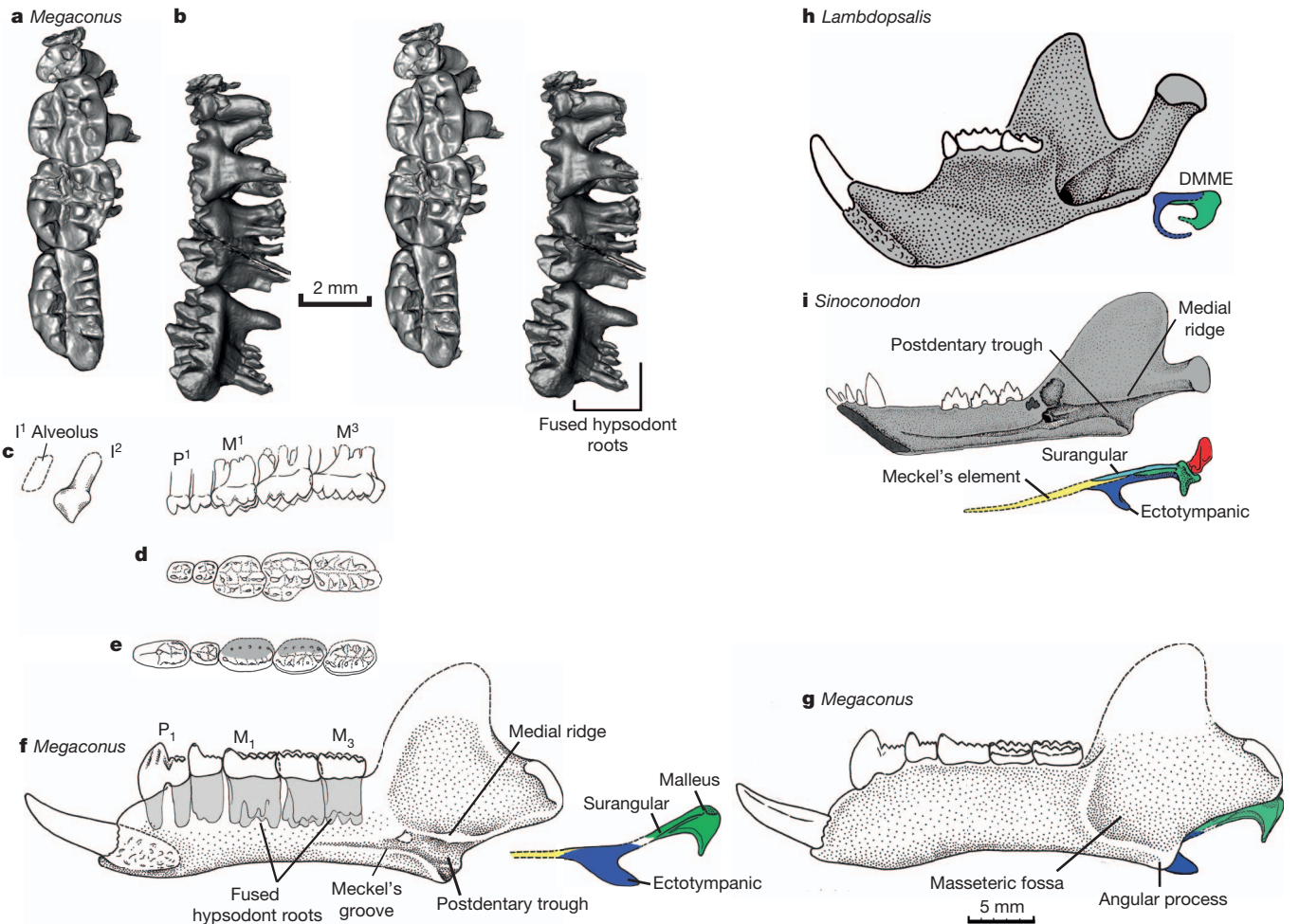
Figure 2 | Dental, mandibular and ear structures of *Megaconus* and comparative taxa. a, b, Occlusal and lingual views of upper P^2–M^3 (stereo images from computed tomography scan, after correction for the fault through M^2; Supplementary Fig. 2 and Supplementary Video 1). **c,** Lingual view reconstruction of I^1 (alveolus only) and I^2–M^3. **d,** Occlusal view of P^1–M^3. **e,** Occlusal view of lower P$_1$–M$_3$ (Supplementary Video 1). **f,** Lingual view of lower teeth (partially fused hypsodont roots in shading), dentary and reconstructed mandibular middle ear. **g,** Labial view reconstruction of mandible with middle ear. **h,** Multituberculate *Lambdopsalis*. **i,** Mandible and middle ear of the mammaliaform *Sinoconodon* (**h, i** from ref. 24). CMME, cynodont mandibular middle ear (ref. 24); DMME, definitive mammalian middle ear. Yellow, ossified Meckel's element ('postdentary rod'); green, malleus (articular) and surangular; blue, ectotympanic (angular); red, incus (quadrate). Details in Supplementary Figs 2–5.
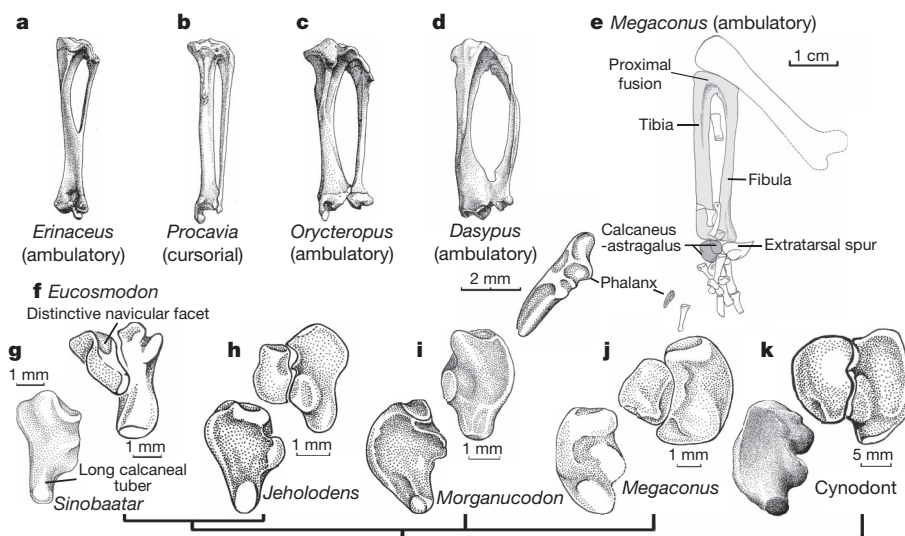
Figure 3 | Comparison of the hindlimb and pes of *Megaconus* and other mammaliaforms. a–d, Fused tibio–fibulas of obligatory terrestrial mammals and *Megaconus*: hedgehog *Erinaceus* (ambulatory terrestrial) (**a**); hyrax *Procavia* (cursorial terrestrial) (**b**); aardvark *Orycteropus* (ambulatory terrestrial) (**c**); armadillo *Dasypus* (ambulatory, terrestrial) (**d**). **a–d** from ref. 37. **e,** *Megaconus*, with fused tibio–fibula (posterior view; fused distal tibio–fibula best seen in Supplementary Fig. 8) and partially disarticulated pes (proximal tarsals in dorsal view). **f, g,** Mediolaterally compressed calcaneal tuber in the multituberculates *Eucosmodon* (dorsal view, from ref. 39) and *Sinobaatar* (ventral view). **h,** Eutriconodont *Jeholodens* (dorsal and ventral views (refs 39, 40)). **i,** Mammaliaform *Morganucodon* (top, dorsal view; bottom, ventral view; redrawn from refs 36, 40). **j,** *Megaconus* (top, based on PMOL-AM00007B; bottom, based on PMOL-AM00007A). **k,** Triassic Manda Formation cynodont (from refs 35, 39).
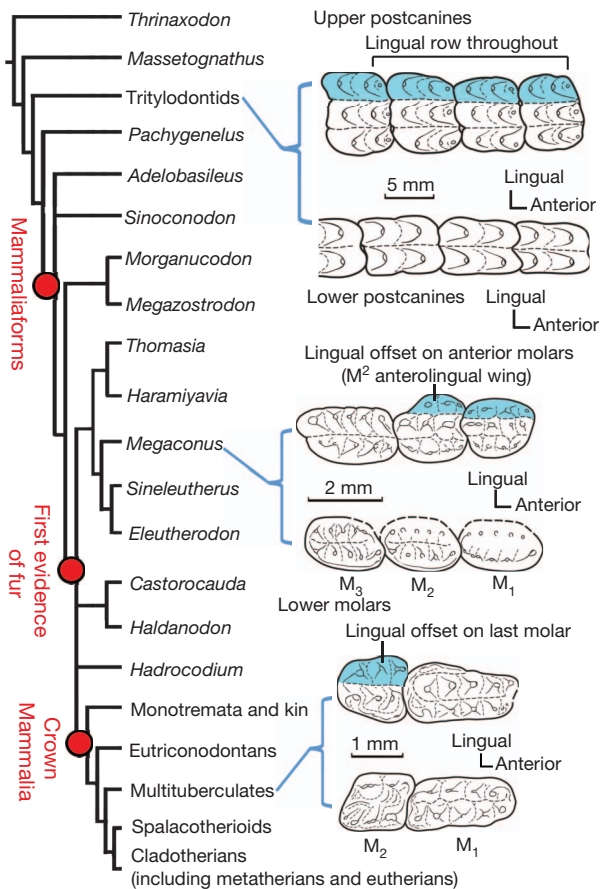
Figure 4 | Cynodont–mammal transition and evolution of mammal-like postcanines with multicusped rows that occlude alternately between uppers and lowers, for omnivory–herbivory feeding adaptations. Eleutherodontids differ from multituberculates in molar counts, and in tooth positions of lingual offset tooth row pattern. *Megaconus* is another case of fur preserved in mammaliaforms, in addition to the case of *Castorocauda*, corroborating that the origins of fur occurred before the origin of crown Mammalia.

The former hypothesis would posit that such complex molariforms evolved, only once, in a combined haramiyidan–multituberculate clade, whereas the latter phylogeny would lead to a scenario of such molariforms evolving twice, first with haramiyidans during the Triassic, and then in multituberculates during the Jurassic. After incorporating numerous features newly revealed by *Megaconus*, in addition to the classic dental features analysed in detail by previous studies[16,26], our phylogeny (Fig. 4 and Supplementary Fig. 9) shows that haramiyidans (including *Megaconus*) are a mammaliaform clade outside of crown Mammalia, separated from multituberculates, which are deeply nested within Mammalia. The hypothesis that suggests that haramiyidans are stem mammaliaforms (Supplementary Fig. 9) is significantly different from the haramiyidan–multituberculate hypothesis by nonparametric test on our matrix ($P \leq 0.00389$). A key implication is that the 'multituberculate-like' molar pattern evolved in haramiyidans in convergence to multituberculates (Fig. 4).

Although molars of haramiyidans and multituberculates show similar alternating occlusion of multicusped rows between the uppers and the lowers in backward power stroke of the mandible[16], we note several obvious differences between these groups: (1) the lower premolars of eleutherodontids (Fig. 2 and Supplementary Fig. 4) have a singular, enlarged and re-curved anterior cusp and a talonid heel that has two short cuspule rows (P_1) or a basin (P_2). These are fundamentally different from the multituberculate lower premolars with crown-length blade of multiple serrations with fine ridges on the sides of the blade. Lower premolars of eleutherodontids are capable of puncturing,

in contrast to the bladed premolars of multituberculates specialized for shearing[29]; (2) basal multituberculates have a full premolar count of five but only two molars[3], whereas eleutherodontids have only two premolars but three molars (Fig. 4); (3) the lingual offset of the last upper molar is the most important synapomorphy of all multituberculates: the lingual cusp row of M^2 is offset from the lingual row of M^1 (refs 3, 16, 30). In *Megaconus*, however, it is M^1 and M^2 that are lingually offset from the last molar M^3 (Fig. 4). The intact dentition scanned by computed tomography (Supplementary Figs 2 and 3) proves that the multiple rows alternately occluding between the upper and the lower molars are developed in different tooth positions in eleutherodontids, as compared to multituberculates; (4) eleutherodontids have multiple, proximally fused hypsodont roots, known only in Cretaceous gondwanatherians among Mesozoic era mammals[3]. No multituberculates have this root pattern; (5) *Megaconus* lacks the anterior extension of the masseteric fossa onto the body of the mandible, as seen in multituberculates for the posteriorly directed power-stroke[30]. *Megaconus* is similar to *Haramiyavia* in this plesiomorphic feature[15], and probably had a much weaker posterior power-stroke of the mandible.

Skeletal features and habits

Megaconus has a gracile postcranial skeleton (Fig. 1 and Supplementary Fig. 1) and is estimated to have weighed about 250 g (Supplementary Information). The vertebral column of *Megaconus* has 24 dorsal vertebrae, several more than those of most Mesozoic mammals (except some eutriconodonts[18,32] and most living mammals). This high dorsal vertebra count is comparable to hyracoids, perissodactyls and some xenarthrans among extant placentals[33]. *Megaconus* and multituberculates differ conspicuously in vertebral features: the thoracolumbar boundary is distinctive and positioned at thoracic 13, and the anticlinal is positioned at thoracic (dorsal) 10 in multituberculates[34]. In *Megaconus*, the thoracolumbar transition is gradational from dorsal vertebrae D16 to D20, the anticlinal vertebra is positioned at D21 (Fig. 1). The anticlinal marks the boundary of the posteriorly inclined neural spines of anterior dorsals, and the anteriorly inclined neural spines of posterior dorsals, corresponding to the regions of the anterior and posterior epaxial muscles. The difference of the anticlinal at D10 in multituberculates versus D21 in *Megaconus* indicates that *Megaconus* had more vertebral segments and a much longer anterior epaxial skeleton–muscular region than multituberculates.

Megaconus is more similar to the Triassic cynodont from the Manda Formation and *Morganucodon*[35,36] than to multituberculates and therian mammals in the calcaneus and astragalus (Fig. 3). The calcaneus has an extensive peroneal shelf, a massive sustentacular region, but a short and ventrally directed calcaneal tuber continuous with the calcaneal body, all of which are similar to cynodonts (Fig. 3j, k)[35] but different from mammals including multituberculates, in which the tuber is distinctive from the main body of the calcaneus (Fig. 3f–i). The astragalus is plesiomorphically similar to cynodonts in the absence of the well-defined navicular facet of most mammaliaforms[36]. An enlarged extratarsal spur, consisting of bony base (os calcares) and horny (keratin) spur (cornu calcares), shows that *Megaconus* had a poisonous spur similar to the living monotremes, now also known from docodont mammaliaforms[7].

The tibia and fibula are fused in *Megaconus* (Fig. 3e). The two bones are fused at both the proximal and the distal ends, as can be verified on both the slab and the counter-slab (Fig. 1 and Supplementary Fig. 1), but not along shafts. Fused tibio–fibula is a common feature of terrestrial mammals (except for the arboreal tarsier)[37]. Cursorial or bipedal jumping (terrestrial) mammals commonly show tibio–fibula fusion along much of the shafts in addition to a great reduction in the fibular part[38]. But mammals with ambulatory (walking) gaits tend to retain an unreduced fibula fully separated along shafts from the tibia. The tibio–fibular fusion occurs only at the proximal and distal ends in *Megaconus* as in armadillos (for example, *Dasypus*) that are terrestrial with ambulatory gaits. The tibio–fibula also bears resemblance to the ambulatory

aardvark (*Orycteropus*) and cursorial hyraxes (*Procavia* and *Heterohyrax*) that have the tibia and fibula fused only at the proximal ends, but remain separated at the distal ends. *Megaconus* lacks an elongate calcaneal tuber, which is an important in-lever for effective action of the Achilles tendon muscles on the ankle, as seen in multituberculates through crown Theria. The hypothesis stating that *Megaconus* is a terrestrial mammaliaform is consistent with the fact that both the manual and pedal terminal phalanges are generalized, with lateral ridges, but without the high and arched dorsal profile as in climbing mammals, or hypertrophied as in digging mammals. The claw keratin sheath preserved as an impression with three claws has a wide outline but only slight dorsoventral curvature (Fig. 1c, 3–5), typical of extant terrestrial mammals.

The *Megaconus* skeleton is preserved with the guard hairs as an impression in most areas of the halo around the skeleton that can be distinguished by the lighter colour of under fur residues. The abdomen appears to have areas of sparse hairs and naked skin with small skin foldings. Thus *Megaconus* probably had a naked abdomen, although it is not possible to speculate whether a pouch would have been present. Furs and other integumentary features (for example, poisonous spur) in eleutherodontids as a mammaliaform clade corroborate the widespread presence of mammalian integument among mammaliaforms, previously known only from a docodontan[7], and that mammalian fur originated earlier in evolutionary history than the ancestor of extant mammals (Fig. 4).

1. Rowe, T. B. Definition, diagnosis, and origin of Mammalia. *J. Vertebr. Paleontol.* **8,** 241–264 (1988).
2. Kermack, K. A., Kermack, D. M., Ires, P. M. & Mills, J. New multituberculate-like teeth from the Middle Jurassic of England. *Acta Palaeontol. Pol.* **43,** 581–606 (1998).
3. Kielan-Jaworowska, Z., Cifelli, R. L. & Luo, Z.-X. *Mammals from the Age of Dinosaurs: Origins, Evolution, and Structure* (Columbia Univ. Press, 2004).
4. Liu, Y.-Q. *et al.* U-Pb zircon age for the Daohugou Biota at Ningcheng of Inner Mongolia and comments on related issues. *Chin. Sci. Bull.* **51,** 2634–2644 (2006).
5. Peng, N., Liu, Y. -Q., Kuang, H. -W., Jiang, X. -J. & Xu, H. Stratigraphy and geochronology of vertebrate fossil-bearing Jurassic strata from Linglongta, Jianchang County, Western Liaoning, Northeastern China. *Acta Geol. Sinica (English Ed.)* **86,** 1326–1339 (2012).
6. Liu, Y.-Q. *et al.* Timing of the earliest known feathered dinosaurs and transitional pterosaurs older than the Jehol Biota. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **323–325,** 1–12 (2012).
7. Ji, Q. *et al.* A swimming mammaliaform from the Middle Jurassic and ecomorphological diversification of early mammals. *Science* **311,** 1123–1127 (2006).
8. Meng, J. *et al.* A Mesozoic gliding mammal from northeastern China. *Nature* **444,** 889–893 (2006).
9. Luo, Z.-X., Ji, Q. & Yuan, C.-X. Convergent dental evolution in pseudotribosphenic and tribosphenic mammals. *Nature* **450,** 93–97 (2007).
10. Luo, Z.-X. *et al.* A Jurassic eutherian mammal and the divergence of marsupials and placentals. *Nature* **476,** 442–445 (2011).
11. von Koenigswald, W. Diversity of hypsodont teeth in mammalian dentitions – construction and classification. *Palaeontographica A* **294,** 63–94 (2011).
12. Martin, T., Averianov, A. O. & Pfretzschner, H.-U. Mammals from the Late Jurassic Qigu Formation in the southern Junggar Basin, Xinjiang, Northwest China. *Palaeodivers. Palaeoenv.* **90,** 295–319 (2010).
13. Averianov, A. O., Lopatin, A. V. & Krasnoluskii, S. A. The first haramiyid (Mammalia, Allotheria) from the Jurassic of Russia. *Dokl. Biol. Sci.* **437,** 103–106 (2011).
14. Sigogneau-Russell, D. Haramiyidae (Mammalia, Allotheria) en provenance du Trias supérieur de Lorraine (France). *Palaeontographica A* **206,** 137–198 (1989).
15. Jenkins, F. A. *et al.* Haramiyids and Triassic mammalian evolution. *Nature* **385,** 715–718 (1997).
16. Butler, P. M. Review of the early allotherian mammals. *Acta Palaeontol. Pol.* **45,** 317–342 (2000).
17. Hahn, G. & Hahn, R. Evolutionary tendencies and systematic arrangement in the Haramiyida (Mammalia). *Geol. Paleontol.* **40,** 173–193 (2006).
18. Luo, Z.-X. *et al.* A new euticonodont mammal and evolutionary development of early mammals. *Nature* **446,** 288–293 (2007).
19. Clemens, W. A. Rhaeto-Liassic mammals from Switzerland and West Germany. *Zitteliana Abh. Bayer. Staatsslg. Paläont. Hist. Geol.* **5,** 51–92 (1980).
20. Hahn, G., Sigogneau-Russell, D. & Wouters, G. New data on Theroteinidae — their relations with Paulchoffatiidae and Haramiyidae. *Geol. Paleontol.* **23,** 205–215 (1989).
21. Kermack, K. A., Mussett, F. & Rigney, H. W. The skull of *Morganucodon*. *Zool. J. Linn. Soc.* **71,** 1–158 (1981).
22. Lillegraven, J. A. & Krusat, G. Cranio-mandibular anatomy of *Haldanodon expectatus* (Docodonta; Mammalia) from the Late Jurassic of Portugal and its implications to the evolution of mammalian characters. *Contrib. Geol. Univ. Wyoming* **28,** 39–138 (1991).
23. Meng, J., Wang, Y.-Q. & Li, C.-K. Transitional mammalian middle ear from a new Cretaceous Jehol euticonodont. *Nature* **472,** 181–185 (2011).
24. Luo, Z.-X. Developmental patterns in Mesozoic evolution of mammal ears. *Annu. Rev. Ecol. Evol. Syst.* **42,** 355–380 (2011).
25. Rougier, G. W., Wible, J. R. & Novacek, M. J. Middle-ear ossicles of *Kryptobataar dashzevegi* (Mammalia, Multituberculata): implications for mammaliamorph relationships and evolution of the auditory apparatus. *Am. Mus. Novit.* **3187,** 1–43 (1996).
26. Ji, Q., Luo, Z.-X., Zhang, X.-L., Yuan, C.-X. & Xu, L. Evolutionary development of the middle ear in Mesozoic therian mammals. *Science* **326,** 278–281 (2009).
27. Jenkins, F. A. Jr & Krause, D. W. Adaptations for climbing in North American multituberculates (Mammalia). *Science* **220,** 712–715 (1983).
28. Crompton, A. W. Postcanine occlusion in cynodonts and tritylodontids. *Bullet. British Mus. (Nat. Hist.) Geol.* **21,** 29–71 (1974).
29. Krause, D. W. Jaw movement, dental function, and diet in the Paleocene multituberculate *Ptilodus*. *Paleobiology* **8,** 265–281 (1982).
30. Lazzari, V. *et al.* Occlusal pattern in paulchoffatiid multituberculates and the evolution of cusp morphology in mammaliamorphs with rodent-like dentitions. *J. Mamm. Evol.* **17,** 177–192 (2010).
31. Butler, P. M. & Hooker, J. J. New teeth of allotherian mammals from the English Bathonian, including the earliest multituberculates. *Acta Palaeontol. Pol.* **50,** 185–207 (2005).
32. Hu, Y.-M. *Postcranial Morphology of* Repenomamus *(Eutriconodonta, Mammalia): Implications for the Higher-Level Phylogeny of Mammals*. PhD thesis, City Univ. New York (2006).
33. Narita, Y. & Kuratani, S. Evolution of the vertebral formulae in mammals: a perspective on developmental constraints. *J. Exp. Zool. B Mol. Dev. Evol.* **304(B),** 91–106 (2005).
34. Kielan-Jaworowska, Z. & Gambaryan, P. P. Postcranial anatomy and habits of Asian multituberculate mammals. *Fossils and Strata* **36,** 1–92 (1994).
35. Jenkins, F. A., Jr. The postcranial skeleton of African cynodonts. *Peabody Mus. Nat. Hist. Bullet.* **36,** 1–216 (1971).
36. Szalay, F. S. *Evolutionary History of the Marsupials and an Analysis of Osteological Characters* (Cambridge Univ. Press, 1994).
37. Lessertisseur, J. & Saban, R. in *Traité de Zoologie Tome XVI (Fascicle I). Mammiferes: Teguments et Skelettes.* (ed. Grassé, P.-P.). 709–1078 (Masson, 1967).
38. Hildebrand, M. & Goslow, G. E. Jr. *Analysis of Vertebrate Structure* 5th edn (John Wiley, 2001).
39. Ji, Q., Luo, Z.-X. & Ji, S.-A. A Chinese triconodont mammal and mosaic evolution of the mammalian skeleton. *Nature* **398,** 326–330 (1999).
40. Luo, Z.-X. & Wible, J. R. A new Late Jurassic digging mammal and early mammalian diversification. *Science* **308,** 103–107 (2005).

Author Contributions C.-F.Z. authenticated the fossil and provided geological data; T.M. organized preparation, CT scan and SEM photography; T.M. and Z.-X.L. performed phylogenetic analyses; S.W., T.M. and Z.-X.L. participated in fossil morphological work and contributed to interpreting the fossil; T.M. and Z.-X.L. developed the manuscript with contributions from all authors.

ARTICLE

Connectomic reconstruction of the inner plexiform layer in the mouse retina

Moritz Helmstaedter[1]†, Kevin L. Briggman[1]†, Srinivas C. Turaga[2]†, Viren Jain[2]†, H. Sebastian Seung[2] & Winfried Denk[1]

Comprehensive high-resolution structural maps are central to functional exploration and understanding in biology. For the nervous system, in which high resolution and large spatial extent are both needed, such maps are scarce as they challenge data acquisition and analysis capabilities. Here we present for the mouse inner plexiform layer—the main computational neuropil region in the mammalian retina—the dense reconstruction of 950 neurons and their mutual contacts. This was achieved by applying a combination of crowd-sourced manual annotation and machine-learning-based volume segmentation to serial block-face electron microscopy data. We characterize a new type of retinal bipolar interneuron and show that we can subdivide a known type based on connectivity. Circuit motifs that emerge from our data indicate a functional mechanism for a known cellular response in a ganglion cell that detects localized motion, and predict that another ganglion cell is motion sensitive.

Information about neuronal wiring has long been the basis of formulating and testing ideas about how computation is performed by neural circuits. Complete[1,2] and partial[3–5] wiring diagrams are being used where available. Whether such diagrams can be created by statistical extrapolation or whether higher-order connectivity is functionally important is highly controversial[6–9]. The assumption that mingling neurites connect (Peters' rule[10]) allows connectivity to be inferred from light-microscopic observations of sparsely stained tissue, but is frequently violated[5,7], showing that connectivity must be explicitly tested rather than inferred from proximity. Simultaneous electrical recordings from several cells can determine and quantify their synaptic connectivity[11–13], but do not allow a comprehensive sampling of connections.

Unlike light microscopy, electron microscopy can follow even the thinnest neurites through densely stained neuropil, and can detect unambiguously whether two cells touch and over which area[1,14]. Serial section transmission electron microscopy was, for example, used to reconstruct the complete wiring diagram of the roundworm *Caenorhabditis elegans*[1,2] and to study synaptic connectivity in the retina[15–17]. Volume electron microscopy data sets hundreds of micrometres in extent[18] have been used to reconstruct—guided by previous functional imaging—specific neural circuits[5,18].

The retina performs a variety of image-processing tasks and is one of the best studied parts of the central nervous system[19]. But, only in few cases, such as for direction sensitivity (reviewed in ref. 20), has the underlying neural computation been plausibly explained, combining information from anatomical studies, electrical recordings and two-photon calcium imaging[5,12,20–23].

Here we combined serial block-face electron microscopy (SBEM)[24] data, crowd-sourced manual annotation[25], machine learning-based boundary detection[26,27], and automatic volume segmentation to reconstruct the neurites of 950 neurons in a 114 μm × 80 μm area of the inner plexiform layer (IPL) and all their contacts in that volume.

We establish the validity of the reconstruction using known circuits and demonstrate its use by classifying cells based on their electron-microscopy-resolution morphology, by isolating a new type of bipolar cell, by showing that the cell-to-type contact area is in some cases tightly controlled and can be used to augment type classification, and by uncovering several cases in which neurite co-stratification does not predict a contact. Among the functional implications of these findings is the prediction that a particular ganglion cell is motion sensitive.

Imaging and reconstruction

We used SBEM because a superior z-resolution[18] and lack of image distortions makes SBEM data sets more easily traced by humans[5] and segmented by computers. The main data set used in this study (e2006) has a volume of more than 1 million μm³, includes all layers that contain intra-retinal synaptic connections, and was stained to enhance plasma-membrane visibility[5], further facilitating traceability and automated segmentation.

Because completely labelling such a volume by hand would be prohibitively expensive (about US$10 million), we tried to establish an entirely automatic reconstruction pipeline. Our SBEM data can be automatically segmented into objects that represent the local cellular geometry accurately[26,27]. But even at voxel error rates of a few per cent, cells get fragmented into many pieces (M.H. *et al.*, manuscript in preparation). Manually traced skeletons, on the other hand, do reliably establish intra-cellular continuity over large distances[25] but allow neither identification nor quantification of cell–cell contacts, and visually inspecting close skeleton encounters[5] is impractical for the large number (roughly 10⁶) expected in our data set. Therefore, we separately created skeletons for all cells by crowd-sourced manual annotation, which is much faster than manual volume tracing[25], and volume segmentation (see below).

Skeletons were created by a team of trained human annotators, which included, over time, more than 224 different students. First, the annotators identified all somata and classified them as photoreceptor ($n > 2,000$), glial ($n = 173$), horizontal ($n = 33$), bipolar ($n = 496$), amacrine ($n = 407$) or ganglion ($n = 47$) cells, based on soma location and emerging neurites (Fig. 1a). Starting from the somata, the annotators skeletonized the neurites of all glial, bipolar (Fig. 1b), amacrine and ganglion cells using the KNOSSOS program[25] (http://www.knossostool.org). Multiple tracings by different annotators (average redundancies: 6,

4 and 4 for ganglion, amacrine and bipolar cells, respectively), were automatically consolidated[25], visually inspected and, in a few cases, manually corrected. A total of >20,000 annotator hours yielded 2.6 m of skeletons, representing 0.64 m of neurite, with estimated[25] error rates of 9, 12 and 6 per ganglion, amacrine and bipolar cell, respectively.

Cell types

We classified all neurons into cell types by visual inspection of the bare skeletons, with a focus on the IPL. We found n = 459 almost complete bipolar cells (Fig. 1c; all reconstructed types and cells are shown in Supplementary Data 1 and 6, respectively). Most bipolar cells clearly belonged to one of the 10 types described previously[28] (Fig. 1c). However, particularly for OFF cone bipolar cells (CBCs) (1–4), some classification ambiguity remained, even after taking into account tiling. A random re-examination of 59 ON CBCs (CBC5–9) found one error.

Seven cells showed no similarity to any of the ten bipolar types[28], but shared a distinct morphology and were designated as XBCs (Fig. 1d and Supplementary Data 2a). XBC axons stratify more narrowly but at the same average depth as CBC5 (Fig. 1d, e). Laterally, XBC axons roam widely, similar to CBC9, but their dendrites are comparatively compact, different from CBC9 (Supplementary data 2b), and their depth suggests that they contact cones.

The dendrites of all ganglion cells and of many amacrine cells extended beyond the data set volume. Many ganglion and amacrine cells could nevertheless be grouped by inspecting their neurites (12 ganglion cell types, Fig. 2a, b; 12 narrow-field amacrine cell types, Fig. 2c, d; 33 medium/wide-field amacrine cell types, including 6 displaced types, Fig. 2e, f and Supplementary Data 1 and 6). We used the type-averaged (for individual variations see Supplementary Data 1) neurite density over depth in the IPL (Fig. 2) to create for all amacrine and ganglion cell types unique identifiers (ac64-73, for example, is an amacrine cell type with first and third quartiles at 64% and 73% IPL depth, respectively). Prominent among cell types previously known (see Supplementary Data 7 for a complete listing) are gc30-63, ac25-31 and ac60-65, corresponding to ON/OFF direction-selective ganglion cells[29] (DSGCs; Fig. 2a) and ON and OFF starburst amacrine cells[16] (SACs; Fig. 2e), respectively.

Contact detection

We next combined the skeletons with an automatic segmentation (Fig. 3), created by first training a convolutional network to detect cell boundaries[27], followed by several growth and merge steps (Fig. 3a). The final volume consolidation into a representation of the cellular geometry was performed by combining for each cell all segments overlapping its skeleton (Fig. 3b, typically several hundred segments; total estimated volume error rate about 3%, see Methods).

Of 1,123 fully volume-reconstructed cells, 173 were glia, 110 were orphans (one-of-a-kind cells or cells without a reasonable neurite morphology), and 840 were the neurons used in the analysis. All contacts (n = 579,724) between them were automatically detected and quantified (Fig. 3c, Supplementary Data 5 and Methods). When testing
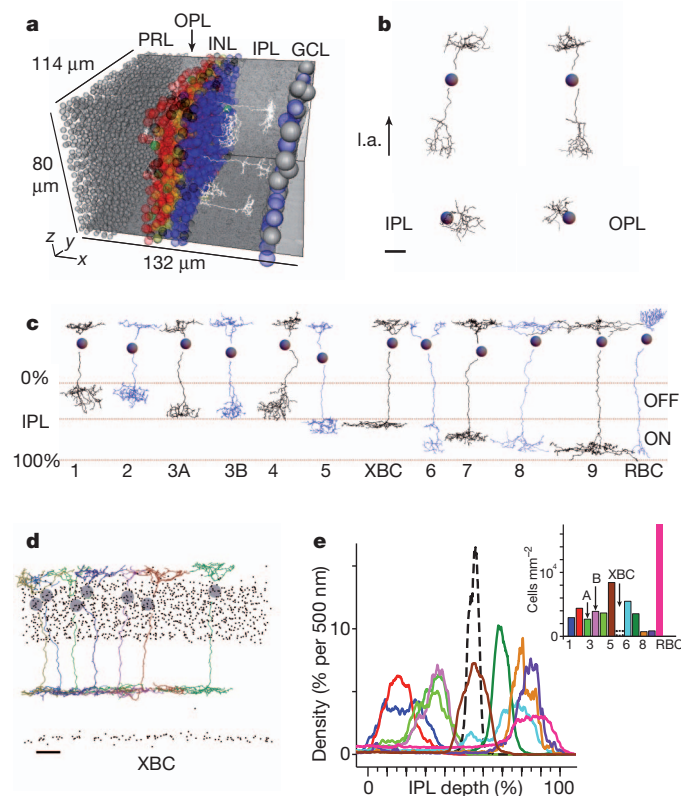
Figure 1 | Raw data, skeletons and bipolar cell analysis. a, Somata, from the left: photoreceptor (grey), horizontal (green), bipolar (red), glia (yellow), amacrine (blue) and ganglion (grey) cells. Also shown (white) are axons for two CBC1, one CBC6 and two CBC7 cells. GCL, ganglion cell layer; INL, inner nuclear layer; OPL, outer plexiform layer; PRL, photoreceptor layer. **b,** Side views from two orthogonal directions onto a single CBC4 skeleton (top), and light-axis (l.a.) views of dendrite (left) and axon (bottom). **c,** One example for each bipolar cell type. **d,** All XBC skeletons, side view. **e,** Skeleton density (segment length/vote count, normalized across IPL) versus depth for all bipolar cell types (one profile shown for the entire CBC5 population). Inset: bipolar cell prevalence (colours as for depth profiles). Scale bars, 10 μm.
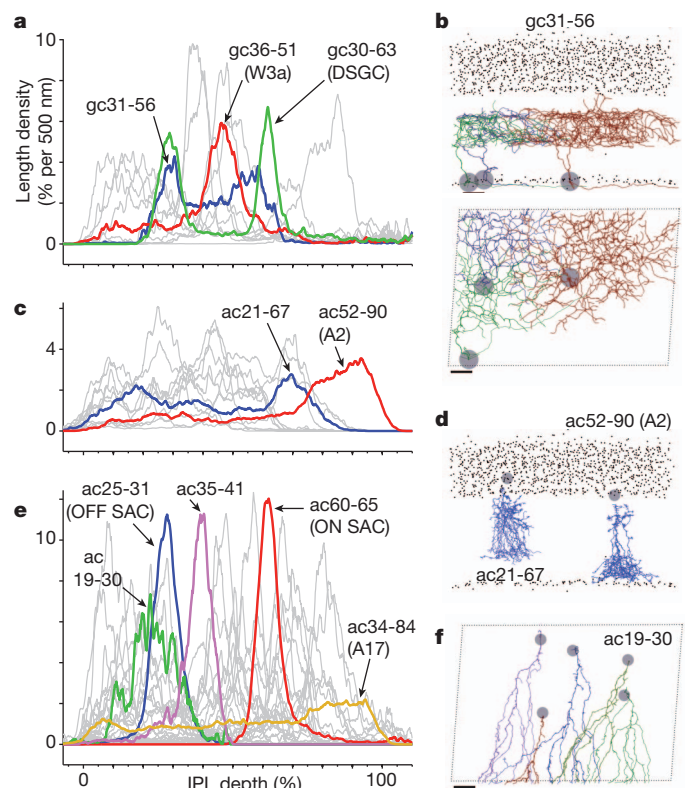
Figure 2 | Ganglion and amacrine cells. a, Normalized ganglion cell depth profiles for gc31-56, gc36-51 (W3), gc30-63 (DSGC) and the remaining cell types (grey). **b,** All three gc31-56 cells (somata: grey disks, side (top) and light-axis (bottom) views), and all other inner nuclear layer and ganglion cell layer (somata: black dots, side view only). **c,** Narrow-field amacrine cell depth profiles for ac21-67, ac52-90 (A2) and remaining narrow-field amacrine cells (grey). **d,** One example each for ac21-67 and ac52-90 (A2). **e,** Medium-field amacrine depth profiles for ac19-30, ac25-31 (OFF SAC), ac35-41, ac60-65 (ON SAC), ac34-84 (A17) and remaining medium-field amacrine cells (grey). **f,** Light-axis view for ac19-30. Scale bars, 10 μm.
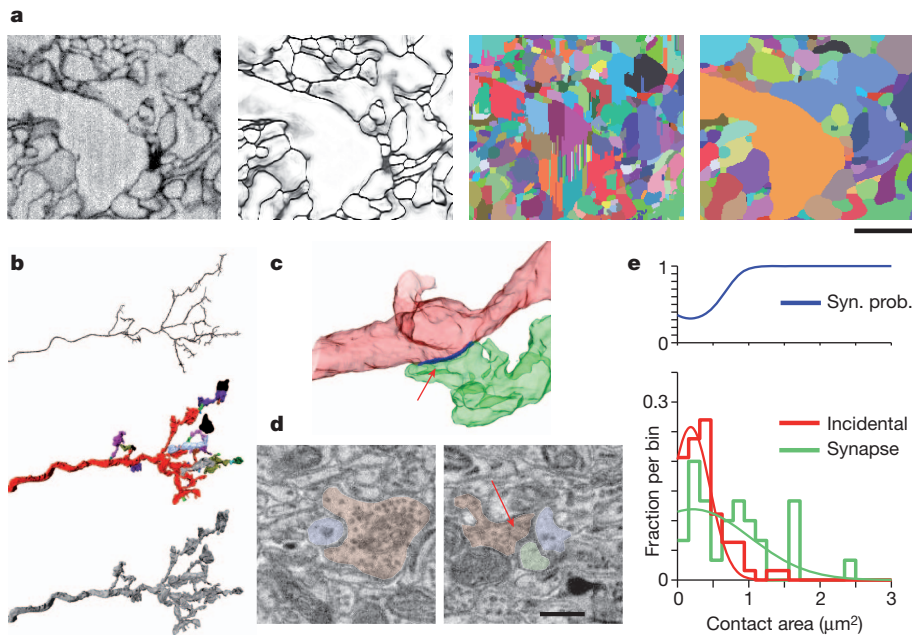
Figure 3 | Automatic segmentation and contact detection. a, From left: raw data (offset and contrast adjusted), edge classifier (*xyz* average), initial and iterated segmentations (see also Supplementary Data 3e–g). **b**, From top: bare skeleton, skeleton with overlapping segmentation objects, and the resulting volume representation for a CBC6 axon. **c**, Automatically detected contact (red arrow) between a CBC5 cell and a DSGC (gc30-63). **d**, Cross-sections through a non-synaptic contact (left, Supplementary Data, 3a, b) and a ribbon synapse (right) from data set k563, coloured by hand. **e**, Frequencies (bottom) of non-synaptic (red, $n = 63$) and synaptic (green, $n = 30$) cell–cell contacts versus contact area and the Gaussian fits to them (thin lines, centre/width: 0.18/0.38 and 0.22/1.13, all in μm^2) and the resulting synapse probability (syn. prob.) estimate (top). Scale bars, 1 μm (**a**), 500 nm (**d**).

the reliability of the algorithm, we found that it missed none of 16 contacts visually identified in the raw data, and 20 randomly selected algorithm-generated contacts contained only one false contact (caused by debris in one image).

The cell-to-cell contact-area matrix (Fig. 4a and Supplementary Data 4) includes only contacts that are individually below 5 μm² (about 99.9% of all contacts), thus excluding touching somata and neurite bundles, and was then condensed into a type-to-type matrix (Fig. 4b and Supplementary Data 4). When exploring the circuit that couples rod photoreceptor signals into the cone pathway[30], we found the A2 amacrine cell (ac52-90) contacting the rod bipolar cell (RBC) very strongly (with 23.6% of the A2 cell total detected neuronal contact area), contacting OFF CBCs quite well (6.4%, 4.3%, 1.8%, 3.2% and 2.9%, for types 1, 2, 3A, 3B and 4, respectively), and contacting ON CBCs more weakly (mostly CBC6 (2.0%) and CBC7 (1.4%), but not the XBC (0.2%)). The RBC, 38.3% of whose contact area is with the A2 cell, also strongly contacts (with 13.5%) ac34-84 (also known as an A17 amacrine cell)[31].

Even when two cell types strongly contact each other (Fig. 4b), the contact area between each individual pair of cells, one from each type, varies widely (Fig. 4a). To test whether individual cells still form reliable channels of information, we compared how the total contact area that a cell of type A makes with all cells of type B varies among the cells of type A. For example, the contact areas between individual ON SACs and all cells of the CBC5R 'type' (9.9% on average, the most strongly contacted one among the CBCs) vary by only about 16% (s.d./mean; Fig. 4c). At the same time, the contact area between A2 amacrine cells (ac52-90) and all RBCs, which is on average even stronger (24%), fluctuates more widely, by 25% (s.d./mean).

To test how much information about the actual synaptic connectivity is provided by our contact-area measurements, we used the size distributions for synaptic and incidental contacts, measured in a data set (k563, ref. 5) with prominently stained synaptic vesicles and thickenings (Fig. 3d, e and Supplementary Data 3a, b), to estimate, for all CBC–ganglion cell pairs, how many true synaptic contacts to expect for a given total contact area between two cells (Fig. 4d), and found that for a total contact area as small as 0.08 μm², at least one synaptic contact exists with a probability of 50%, increasing to 95% for an area of 1 μm².

Connectivity-based type classification

We next explored whether comprehensive contact information contained in the cell-to-cell matrix can be used to discriminate between otherwise very similar cell types. When we searched for a way to divide the CBC5s, which fall into two molecularly distinguishable classes in rats[32] and are too numerous for a single class in mouse[28], by using their connectivities to ganglion cells and amacrine cells, gc31-56 and gc36-51 emerged as potential discriminators (Fig. 5a). A reasonably complete tiling pattern resulted (Fig. 5b) when including only cells ($n = 22$) contacting gc31-56 more strongly than gc36-51 (the exception was a single cell, which was near that threshold but was not included to avoid strong axonal overlap; asterisk in Fig. 5a). This group of cells, 'CBC5A', also shows a strong repulsion between their dendritic centroids (Fig. 5c), indicating a mosaic and hence a pure type[33], and is specifically avoided by ac43-49 (Fig. 5a). The remaining 37 cells ('CBC5R') still show strong axonal overlap, lack a mosaic (Fig. 5b, c), and are thus probably a mixture of types for which we did not, however, find a connectivity-based discriminator. The depth profiles of CBC5A (first and third quartiles: 54% and 61%) and CBC5R (50% and 59%; Fig. 5d) seem to be different. Ten cells did not overlap the dendrites of both ganglion cell types (Fig. 5b) and were therefore collected into a separate group ('CBC5X').

XBC circuits

We next investigated how the XBC is integrated into the IPL circuitry (Fig. 6a–c). Like RBC and CBC7, XBC devotes less of its contact area to ganglion cells than the average bipolar cell (Fig. 6a). XBC strongly contacts (Supplementary Data 2b) medium/wide-field amacrine cells ac38-56 (15.5%) and ac53-59 (7.1%), of which ac53-59 shares the XBC sharp depth profile (Fig. 6b) and, in turn, makes contact with gc31-56 (3.5%) and gc47-57 (4.2%). Those ganglion cells, however, receive only minimal amounts (0.9% and 0.4%) of their contacts directly from the XBC, even though their dendrites strongly overlap XBC axons in depth (Fig. 6b). Instead gc31-56 receives direct bipolar cell contacts mainly from CBC5A (7.0%) and gc47-57 from CBC5R (12.0%). ac38-56 is bistratified, overlapping in the ON stratum with the XBC and in the OFF stratum with gc35-41 (Fig. 6b, c), which is clearly an OFF cell (contacting CBCs 3A, 3B and 4, with 5.4%, 6.3% and 5.4%, respectively; all other CBCs are at most 0.5%) and receives 10.0% of its contacts from ac38-56.

ON/OFF ganglion cell circuits

Some of the best studied ganglion cells respond to both ON and OFF stimuli. We therefore analysed the connection patterns onto several
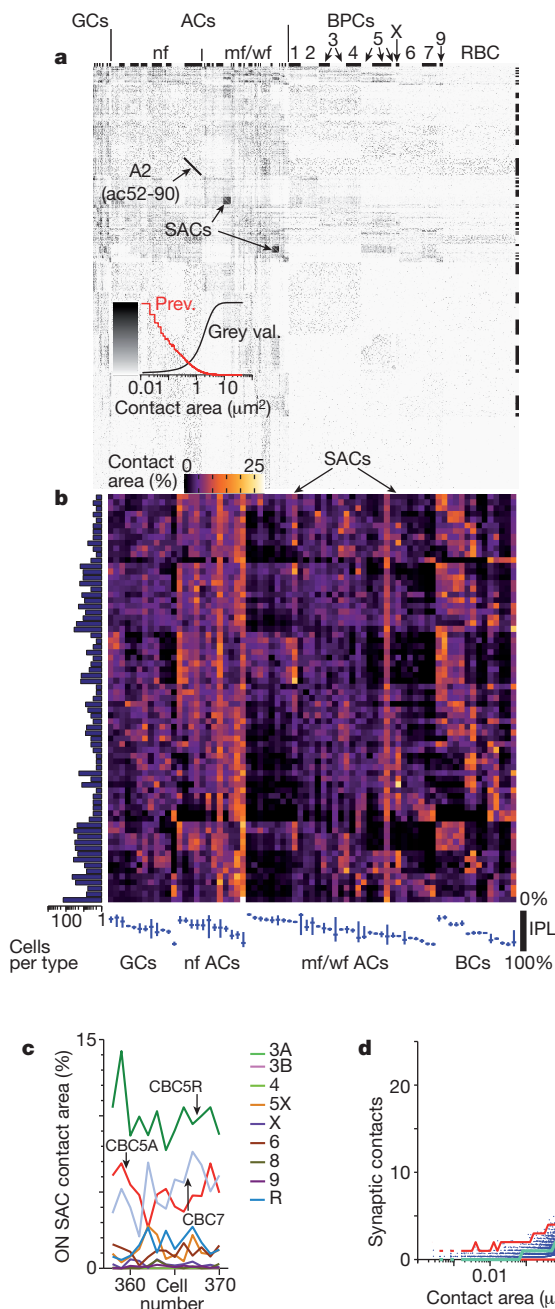
Figure 4 | Contact matrices. a, Cell–cell contact-size matrix (see also Supplementary Data 4 and 5). Classes: ganglion cells (GCs), amacrine cells (ACs), bipolar cells (BPCs), from left to right. Ordering within classes: ganglion cells, types by depth (average of first and third quartile) in the IPL. Amacrine cells: narrow-field (nf), medium/wide-field (mf/wf) (including displaced). Within sub-classes: by depth, except bipolar cells (by numbering in ref. 28, XBC after 5X, and RBC last). Within types: random order. Dark lines along top and right side: every other type. Inset (bottom left): contact area-to-grey value (grey val.) mapping and prevalence (prev.). **b**, Type–type matrix, normalized along rows; along left edge: cells/type (note log scale); along bottom edge: median and depth range (between quartiles) for each type. **c**, Total contact area between each ON SAC (ac60-65, cell numbers along the *x* axis) and, respectively, all CBCs and RBCs, normalized to the total contact area of each SAC. **d**, The number of true synaptic contacts expected (Stochastic simulation, 1,000 runs per cell pair) for each actual CBC-to-ganglion-cell pair using the fits in Fig. 3e versus the total cell–cell contact area (median: green line, lower and upper 95% confidence levels, red lines).

ganglion cells that ramify in both ON and OFF layers (Fig. 6d–f). Among those, gc36-51 ('W3a') and gc44-52 ('W3b') are consistent with cells labelled in the TYW3 mouse[34]. Either or both are likely to be
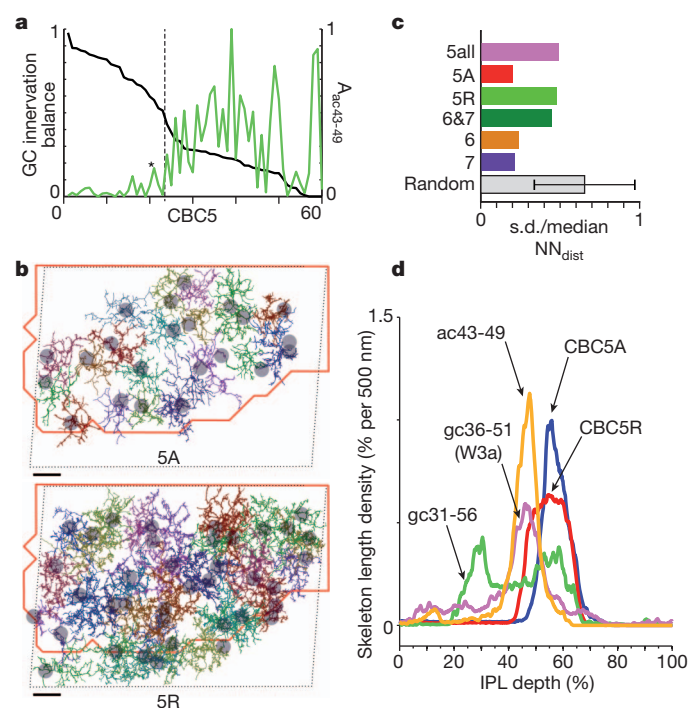
Figure 5 | CBC5 subtypes. a, CBC5s ordered by decreasing preference for gc31-56 over gc36-51 (black trace: contact area with all gc31-56s/sum of contact areas to gc31-56 and gc36-51). Also shown is the contact area with ac43-49 (A_{ac43-49}, green trace, relative to max). Dashed line denotes border between CBC5A and CBC5R. Asterisk denotes CBC5 cell switched to CBC5R to avoid mosaic violation. **b**, Light-axis views of CBC5A (5A; top) and CBC5R (5R; bottom) axons. Red outline: region containing dendrites from both gc36-51 and gc31-56; thin dashed line: data set border. Scale bars, 10 μm. **c**, Variation of dendritic–centroid nearest-neighbour distances (NN_{dist}) (standard deviation/median) for: all CBC5s, only 5A, only 5R, the mixture of 6 and 7, only 6, only 7, and a set of 1,000 simulations randomly placing 22 points (error bar denotes fifth to ninety-fifth percentile). **d**, Normalized skeleton density depth profiles.

homologous to what is called the 'local edge detector' in rabbit[35,36] (Fig. 6d). Their contact patterns with CBCs are mostly similar (gc36-51/gc44-52: CBC5R, 7.5%/11.5%; CBC5A, 1.3%/0.8%; CBC4, 3.0%/3.9%; CBC3A, 1.7%/1.8%; and CB3B, 3.2%/1.7%; Supplementary Data 1), with the exception of the outermost part of the inner nuclear layer (INL) (CBC2, 1.5%/0.1%, and CBC1, 1.6%/0.1%). Substantial contacts are made by gc36-51 and gc44-52 with several narrow-field amacrine cells, ac52-90 (6.0%/2.8% (ref. 37), A2), ac21-67 (3.8%/2.1%), ac51-70 (3.5%/5.0%) and ac21-44 (3.3%/2.2%). The strongest amacrine cell contact made by gc36-51 is with ac43-49 (6.8%), which straddles the boundary between ON and OFF layers (Supplementary Data 1), and also substantially contacts gc44-52 (5.6%) as well as ON and OFF bipolar cells (CBC5R, 9.3%, and CBC4, 5.0%). ac43-49 is one of two medium/wide-field amacrine cells that dedicate most of their contacts to gc36-51 and gc44-52 (Supplementary Data 1). The second is ac44-54 (7.0%/6.2%), a cell dominated by ON CBCs (7.9% with CBC5R compared to 1.3%, 2.0% and 1.2%, with CBCs 3A, 3B and 4, respectively).

The ON/OFF DSGC (gc30-63, Fig. 6f), as expected[5,21], strongly contacts SACs (9.2% and 11.4%, for ac25-31 (OFF SAC) and ac60-65 (ON SAC)), but substantial contacts from other medium/wide-field amacrine cells are conspicuously absent (ac34-84, 2.5%, all others < 1.6%). Like gc36-51/gc44-52 (W3a/b), the DSGC prefers CBC5R (6.9%) to CBC5A (1.9%, all other ON CBCs at most 1.1%). Its main OFF 'input' comes from CBC4 (3.2%) and CBC3 (A/B, 3.0%/2.7%). SACs make most contacts (Fig. 6e) among themselves (26.6% and 21.4% for ON and OFF). They discriminate less than the DSGC between CBC5R (9.7%) and CBC5A (5.0%), but, most notably, contact
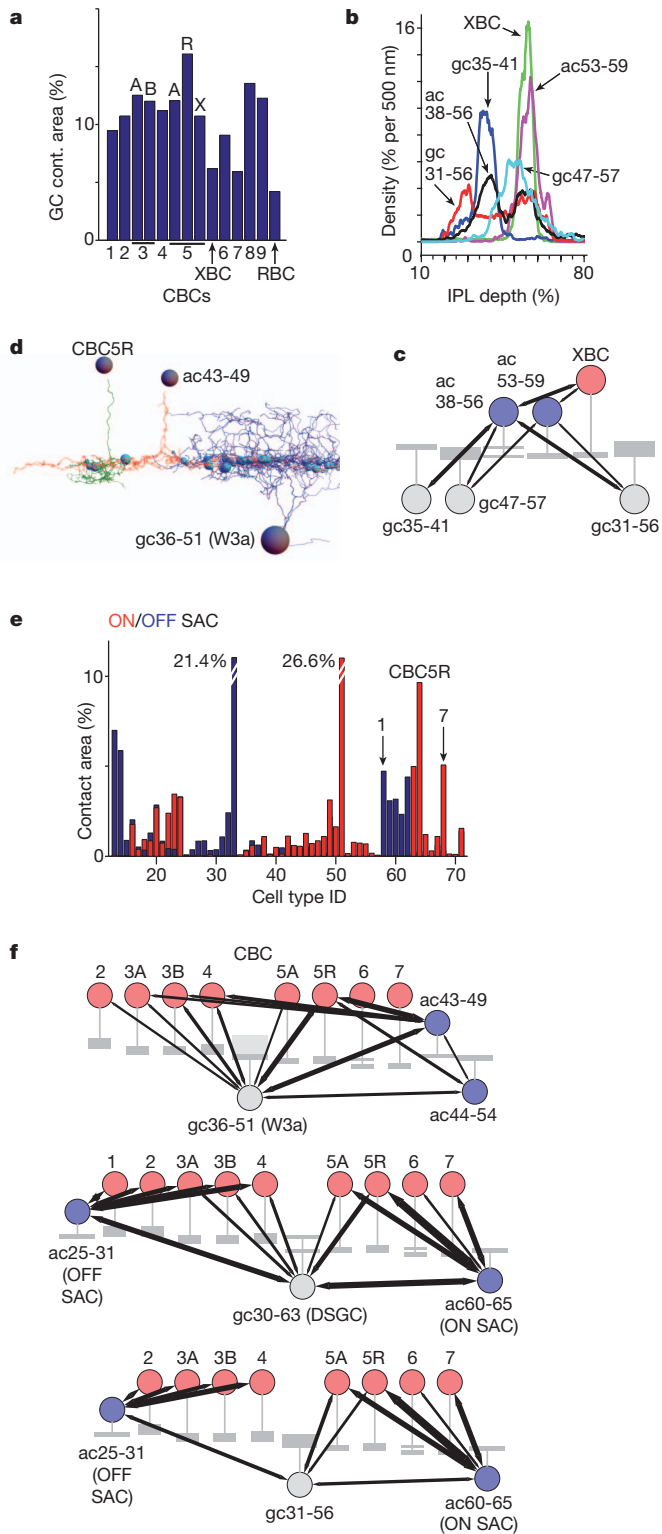
Figure 6 | Circuits originating from the XBC and ON/OFF cells. a–f, IPL circuitry from the XBC (**a–c**) and from three ON/OFF cells (**d–f**). **a**, Fractional contact areas between all ganglion cells and each bipolar cell. **b**, Depth profiles. **c**, XBC circuit schematic. **d**, One gc36-51/W3a (cell 16), one ac43-49 (cell 307), one CBC5R (cell 578) and all the detected contacts between (cyan spheres, volume proportional to contact area). **e**, Normalized contact areas for amacrine cells and bipolar cells with both SACs (ac25-31, blue; ac60-65, red). **f**, Circuit diagrams. Arrow width in circuit diagrams proportional to (total contact area between types)$^{0.5}$. Only connections with areas per type >30 μm^2 shown.

CBC7 (5.1%), which is largely ignored by the DSGC (1.1%, Fig. 6f). Similar differences are seen for the OFF sublamina: DSGC and SAC contact strengths to CBC1/CBC2 are 1.4%/0.5% and 4.7%/3.1%, respectively.

Our last example is the analysis of a cell not associated with any known type in mouse but possibly homologous to a rabbit retina[38] ON/OFF ganglion cell. gc31-56 is an ON/OFF cell by stratification (Fig. 2a, b), filling the space between the SAC bands (Fig. 2a, e), and 'connects' strongly to both SACs (ac60-65, 5.4%, and ac25-31, 7.1%). Surprising is the strong imbalance between ON and OFF bipolar cell 'input' (7.0%/3.7% for CBC5A/R, but only 0.8%, 0.7%, 0.9%, 1.5% and 1.2%, for CBC1, 2, 3A, 3B and 4).

Discussion

Our comprehensive analysis of the bipolar cells confirmed the existence of the ten bipolar cell types previously identified[28], and revealed the existence of the XBC, which had not emerged even in large genetic screens[39]. Although sharp stratification and large size (Fig. 1d, e and Supplementary Data 2a, also note the similarity to cluster 6 in ref. 40) suggest homology between the XBC and the giant bipolar cell described recently in the primate retina[41], the small size of the XBC dendrites relative to its axonal arbour argues against it. The functional role of the XBC is unclear. Its sparseness suggests low spatial resolution and its small dendritic fields suggest that it does not collect signals from all cells of one cone type, thus potentially forgoing some amount of signal. Curious is the absence of a bipolar cell with a similarly sharp stratification on the OFF side. Instead, we find an inter-layer connection via the symmetrically bistratified ac38-56 (Fig. 6b, c). One might speculate that the XBC is part of a luminance adaptation pathway.

Dense sampling and the complete high-resolution reconstruction of neurites, as is only possible with three-dimensional electron microscopy data, contributes in several ways to cell-type classification. First, when all cells of a class, for example, all bipolar cells, are reconstructed, no type will be missed and the prevalence of different types can be determined precisely (Fig. 1e, inset). Second, differences in neurite geometry can be compared for cells within the same piece of tissue. For almost all bipolar cells and a substantial fraction of ganglion and amacrine cells, it was thus possible to establish a correspondence to cell types described in the literature (Supplementary Data 7). We generally erred on the side of splitting groups and expect that some groups actually belong to the same type (for example, the similar connectivity to the XBC suggests that ac38-56 and ac37-52 could be the same type; Supplementary Data. 4). Third, even if they cannot be selectively stained and imaged, tiling and mosaic formation (both used to assess purity of type[33]) can be easily assessed (Figs 2b and 5b and Supplementary Data 1). Fourth, complete contact information can confirm or refine the definition of types (Fig. 5), and may ultimately become sufficient for classification all by itself[42].

Because of the constrained size of our data set, many amacrine cell and all ganglion cell neurites are truncated, and many larger neuron types are presumably completely missed[31,43]. Advances in volume electron microscopy technology[18] now make it possible to acquire volumes with a lateral extent of at least 500 μm. One might then, using the same tools and a similar manual annotation effort as were used in our study, densely reconstruct a central region of 100 μm in extent and trace neurites of passage far enough into the periphery to determine their cell type.

Although our analysis provides contact areas and not synaptic strength, the absence of contact always indicates a lack of synaptic connection. The absence of contacts between some cell types, for example, XBC and gc35-41 as well as CBC7 and DSGC, the neurites of which mingle extensively, confirms that Peters' rule[10] is routinely violated. Furthermore, it seems that large contacts are quite likely to be synaptic (at least between bipolar cells and ganglion cells; Figs 3e and 4d). Although we have not used them here, other geometric parameters describing contact shape might provide enough additional

information to identify actual contacts with near certainty for many types of synapses.

It has been our consistent experience that selectively enhancing cell-surface contrast[5] simplifies manual tracing and enables automatic volume segmentation. If recent results that suggest that even conventionally stained tissue can be reliably traced by hand (K.L.B. and M.H., unpublished observations) and automatically segmented[44] (M. Berning and M.H., personal communication) are confirmed it may no longer be necessary to trade traceability for synapse identification.

The reliability of the entries in the contact matrix depends on several factors. Likely dominant are neurite-continuity errors, which occur roughly six times per bipolar cell[25] but presumably mostly in the periphery and thus should cause only a small fractional loss (or false addition) of synapses. Local volume reconstruction seems to be fairly reliable. Finally, although not all contacts are synaptic, there are, typically, many contacts between any actually connected pair of cells[45,46], making it unlikely that any strong connection is spurious. The connectivity estimate between CBC5R (38 cells) and W3a (gc36-51, 3 cells), for example, is based on the areas of 1,358 observed contacts, for which our simulation predicts between 278 and 705 synaptic contacts (fifth and ninety-fifth percentiles, respectively) with a median of 483 contacts, that is, 13 per bipolar cell and 161 per ganglion cell. The direction of a potential synaptic connection can in most cases not be determined by visually inspecting the e2006 data set but contacts onto a ganglion cell, for example, are presumably never postsynaptic.

Our analysis of three ON/OFF layer cell types has several concrete functional implications, which, at the very least, will guide further exploration by other means. For example, 'bright' W3 cells (gc36-51 or gc44-52) respond much more vigorously to a small darkening spot than to a small brightening spot (Fig. 5b in ref. 47). This may well be due to the ac44-54-mediated feed-forward inhibitory pathway on the ON side (Fig. 6f and Supplementary Data 1), for which no corresponding pathways is seen on the OFF side. W3b–CBC contacts are concentrated on the ON side (ON/OFF: 15.2%/7.6%) but evenly balanced for W3a (10.9%/10.9%), which suggests that it is W3a (gc36-51) that corresponds to the physiologically examined cells described previously[47]. Another characteristic of the W3 cell is that its response is completely suppressed by movement in the receptive-field surround[48]. Given the lack of thin, unbranched processes emerging from its soma, it is unlikely that ac43-49 corresponds to the poly-axonal amacrine cell implicated in this suppression[48], but ac43-49 may well mediate (or at least augment) suppression for stimuli in the near-surround (M. Meister, personal communication). It could do this for OFF and for ON stimuli because it is contacted by CBCs in both OFF and ON layers (Fig. 6f and Supplementary Data 1).

In addition to DSGCs (gc30-63), SACs contact gc31-56 strongly (Fig. 6f). It will be interesting to find out whether gc31-56 is also direction sensitive, or at least motion sensitive, and why there is a morphological symmetry (Fig. 2b) between the ON and OFF layers in gc31-56 but a strong imbalance between the strong ON bipolar cell and the weak OFF bipolar cell 'input' (Fig. 6f).

The circuit motifs found for W3a/b, XBC and gc31-56 are only the first of many examples of motifs likely to be found when these data (a repository of raw data, skeletons and volume segmentation can be found at http://www.neuro.mpg.de/connectomics) are examined in the context of virtually every functional question in the retina.

METHODS SUMMARY

Tissue preparation for SBEM. The retinae for the e2006 and k563 data sets were prepared as described previously[5].

SBEM imaging and data analysis. The sample was mounted in a custom-built ultra-microtome operating inside the chamber of a field-emission scanning electron microscope (FEI QuantaFEG 200), and serial block-face imaged under 130 Pa hydrogen, at 3 keV landing energy, a dose of 14 electrons per nm^2, and a resolution of $16.5 \times 16.5 \times 25\ nm^3$ (for the conventionally stained sample, see Methods). A custom-designed back scattered-electron detector was used. SBEM data were aligned and stitched using custom Matlab routines. Skeletons were manually traced by trained student annotators using custom written software (KNOSSOS, http://www.knossostool.org) and consolidated using RESCOP[25]. Volumes were traced using KLEE (M.H. et al., manuscript in preparation). Boundary classification was with a five-hidden-layer convolutional neural network that was trained using the MALIS procedure[49] (S.C.T. et al., manuscript in preparation). Segmentation used a 15-step iterative growth procedure, followed by a 6-step merging procedure. Data visualization was in KLEE, Knossos, Matlab, Mathematica and Amira.

Full Methods and any associated references are available in the online version of the paper.

1. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans. Phil. Trans. R. Soc. Lond. B* **314,** 1–340 (1986).
2. Varshney, L. R., Chen, B. L., Paniagua, E., Hall, D. H. & Chklovskii, D. B. Structural properties of the *Caenorhabditis elegans* neuronal network. *PLOS Comput. Biol.* **7,** e1001066 (2011).
3. Binzegger, T., Douglas, R. J. & Martin, K. A. C. A quantitative map of the circuit of cat primary visual cortex. *J. Neurosci.* **24,** 8441–8453 (2004).
4. Helmstaedter, M., de Kock, C. P., Feldmeyer, D., Bruno, R. M. & Sakmann, B. Reconstruction of an average cortical column *in silico. Brain Res. Rev.* **55,** 193–203 (2007).
5. Briggman, K. L., Helmstaedter, M. & Denk, W. Wiring specificity in the direction-selectivity circuit of the retina. *Nature* **471,** 183–188 (2011).
6. Stepanyants, A. & Chklovskii, D. B. Neurogeometry and potential synaptic connectivity. *Trends Neurosci.* **28,** 387–394 (2005).
7. Mishchenko, Y. et al. Ultrastructural analysis of hippocampal neuropil from the connectomics perspective. *Neuron* **67,** 1009–1020 (2010).
8. Hill, S. L., Wang, Y., Riachi, I., Schurmann, F. & Markram, H. Statistical connectivity provides a sufficient foundation for specific functional connectivity in neocortical neural microcircuits. *Proc. Natl Acad. Sci. USA* **109,** E2885–E2894 (2012).
9. Denk, W., Briggman, K. L. & Helmstaedter, M. Structural neurobiology: missing link to a mechanistic understanding of neural computation. *Nature Rev. Neurosci.* **13,** 351–358 (2012).
10. Peters, A. Thalamic input to the cerebral cortex. *Trends Neurosci.* **2,** 183–185 (1979).
11. Markram, H., Lubke, J., Frotscher, M., Roth, A. & Sakmann, B. Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. *J. Physiol. (Lond.)* **500,** 409–440 (1997).
12. Fried, S. I., Munch, T. A. & Werblin, F. S. Mechanisms and circuitry underlying directional selectivity in the retina. *Nature* **420,** 411–414 (2002).
13. Asari, H. & Meister, M. Divergence of visual channels in the inner retina. *Nature Neurosci.* **15,** 1581–1589 (2012).
14. Stevens, J. K., Davis, T. L., Friedman, N. & Sterling, P. A systematic approach to reconstructing microcircuitry by electron microscopy of serial sections. *Brain Res.* **2,** 265–293 (1980).
15. Sterling, P. Microcircuitry of the cat retina. *Annu. Rev. Neurosci.* **6,** 149–185 (1983).
16. Famiglietti, E. V. Synaptic organization of starburst amacrine cells in rabbit retina: analysis of serial thin sections by electron microscopy and graphic reconstruction. *J. Comp. Neurol.* **309,** 40–70 (1991).
17. McGuire, B. A., Stevens, J. K. & Sterling, P. Microcircuitry of bipolar cells in cat retina. *J. Neurosci.* **4,** 2920–2938 (1984).
18. Briggman, K. L. & Bock, D. D. Volume electron microscopy for neuronal circuit reconstruction. *Curr. Opin. Neurobiol.* **22,** 154–161 (2012).
19. Masland, R. H. The neuronal organization of the retina. *Neuron* **76,** 266–280 (2012).
20. Vaney, D. I., Sivyer, B. & Taylor, W. R. Direction selectivity in the retina: symmetry and asymmetry in structure and function. *Nature Rev. Neurosci.* **13,** 194–208 (2012).
21. Euler, T., Detwiler, P. B. & Denk, W. Directionally selective calcium signals in dendrites of starburst amacrine cells. *Nature* **418,** 845–852 (2002).
22. Zhou, Z. J. & Lee, S. Synaptic physiology of direction selectivity in the retina. *J. Physiol. (Lond.)* **586,** 4371–4376 (2008).
23. Wei, W., Hamby, A. M., Zhou, K. & Feller, M. B. Development of asymmetric inhibition underlying direction selectivity in the retina. *Nature* **469,** 402–406 (2011).
24. Denk, W. & Horstmann, H. Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS Biol.* **2,** e329 (2004).
25. Helmstaedter, M., Briggman, K. L. & Denk, W. High-accuracy neurite reconstruction for high-throughput neuroanatomy. *Nature Neurosci.* **14,** 1081–1088 (2011).
26. Jain, V. et al. Supervised learning of image restoration with convolutional networks. *IEEE 11th International Conference on Computer Vision* **2,** 1–8 (2007).
27. Turaga, S. C. et al. Convolutional networks can learn to generate affinity graphs for image segmentation. *Neural Comput.* **22,** 511–538 (2010).
28. Wässle, H., Puller, C., Muller, F. & Haverkamp, S. Cone contacts, mosaics, and territories of bipolar cells in the mouse retina. *J. Neurosci.* **29,** 106–117 (2009).
29. Amthor, F. R., Oyster, C. W. & Takahashi, E. S. Morphology of on-off direction-selective ganglion cells in the rabbit retina. *Brain Res.* **298,** 187–190 (1984).
30. Strettoi, E., Raviola, E. & Dacheux, R. F. Synaptic connections of the narrow-field, bistratified rod amacrine cell (AII) in the rabbit retina. *J. Comp. Neurol.* **325,** 152–168 (1992).

31. Macneil, M. A., Heussy, J. K., Dacheux, R. F., Raviola, E. & Masland, R. H. The shapes and numbers of amacrine cells: matching of photofilled with Golgi-stained cells in the rabbit retina and comparison with other mammalian species. *J. Comp. Neurol.* **413,** 305–326 (1999).

32. Fyk-Kolodziej, B. & Pourcho, R. G. Differential distribution of hyperpolarization-activated and cyclic nucleotide-gated channels in cone bipolar cells of the rat retina. *J. Comp. Neurol.* **501,** 891–903 (2007).

33. Wassle, H. & Riemann, H. J. Mosaic of nerve-cells in mammalian retina. *Proc. R. Soc. Lond. B* **200,** 441–461 (1978).

34. Kim, I. J., Zhang, Y., Meister, M. & Sanes, J. R. Laminar restriction of retinal ganglion cell dendrites and axons: subtype-specific developmental patterns revealed with transgenic markers. *J. Neurosci.* **30,** 1452–1462 (2010).

35. Levick, W. R. Receptive fields and trigger features of ganglion cells in the visual streak of the rabbits retina. *J. Physiol. (Lond.)* **188,** 285–307 (1967).

36. Amthor, F. R., Takahashi, E. S. & Oyster, C. W. Morphologies of rabbit retinal ganglion cells with complex receptive fields. *J. Comp. Neurol.* **280,** 97–121 (1989).

37. Kolb, H., Nelson, R. & Mariani, A. Amacrine cells, bipolar cells and ganglion cells of the cat retina: a Golgi study. *Vision Res.* **21,** 1081–1114 (1981).

38. Sivyer, B., Venkataramani, S., Taylor, W. R. & Vaney, D. I. A novel type of complex ganglion cell in rabbit retina. *J. Comp. Neurol.* **519,** 3128–3138 (2011).

39. Siegert, S. *et al.* Genetic address book for retinal cell types. *Nature Neurosci.* **12,** 1197–1204 (2009).

40. Badea, T. C. & Nathans, J. Quantitative analysis of neuronal morphologies in the mouse retina visualized by using a genetically directed reporter. *J. Comp. Neurol.* **480,** 331–351 (2004).

41. Joo, H. R., Peterson, B. B., Haun, T. J. & Dacey, D. M. Characterization of a novel large-field cone bipolar cell type in the primate retina: evidence for selective cone connections. *Vis. Neurosci.* **28,** 29–37 (2011).

42. Seung, H. S. Reading the book of memory: sparse sampling versus dense mapping of connectomes. *Neuron* **62,** 17–29 (2009).

43. Masland, R. H. The fundamental plan of the retina. *Nature Neurosci.* **4,** 877–886 (2001).

44. Andres, B. *et al.* in *Computer Vision – ECCV 2012 Lecture Notes in Computer Science* (eds Fitzgibbon, A. *et al.*) 778–791 (Springer, 2012).

45. Tsukamoto, Y., Morigiwa, K., Ueda, M. & Sterling, P. Microcircuits for night vision in mouse retina. *J. Neurosci.* **21,** 8616–8623 (2001).

46. Calkins, D. J. & Sterling, P. Microcircuitry for two types of achromatic ganglion cell in primate fovea. *J. Neurosci.* **27,** 2646–2653 (2007).

47. Zhang, Y., Kim, I. J., Sanes, J. R. & Meister, M. The most numerous ganglion cell type of the mouse retina is a selective feature detector. *Proc. Natl Acad. Sci. USA* **109,** E2391–E2398 (2012).

48. Ölveczky, B. P., Baccus, S. A. & Meister, M. Segregation of object and background motion in the retina. *Nature* **423,** 401–408 (2003).

49. Turaga, S. C., Briggman, K., Helmstaedter, M., Denk, W. & Seung, H. S. Maximin affinity learning of image segmentation. *Adv. Neural Info. Proc. Syst.* **22,** 1–8 (2009).

Supplementary Information is available in the online version of the paper.

Author Contributions M.H. and W.D. designed the study. K.L.B. prepared the samples and acquired the data using a microtome designed by W.D. M.H. analysed the data, with minor contributions from W.D. S.C.T., V.J. and H.S.S. developed the boundary classifier. M.H., K.L.B. and W.D. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.H. (mhelmstaedter@neuro.mpg.de).

METHODS

Data acquisition. A retina from a 30-day old-C57BL/6 mouse (data set e2006) was prepared to selectively enhance cell outlines by using the horseradish peroxidase (HRP)-mediated precipitation of 3,3'-diaminobenzidine (DAB), as described previously[5], and stained with osmium and lead citrate. The shrinkage of our tissue was very likely the same as that for the e2198 sample[5], which was imaged in the living state by two-photon microscopy and then by SBEM, allowing a precise estimate (14%) of the linear shrinkage factor (K.L.B. *et al.*, unpublished observations). All procedures were approved by the local animal care committee and were in accordance with the law of animal experimentation issued by the German Federal Government.

The embedded tissue was trimmed to a block face of $\sim$200 μm $\times$ 300 μm and imaged in a scanning electron microscope with a field-emission cathode (QuantaFEG 200, FEI Company) and a custom-designed back scattered-electron detector based on a silicon diode (AXUV, International Radiation Detectors) combined with a custom-built current amplifier (J. Tritthardt, Max Planck Institute for Medical Research, electronics shop). The incident-electron energy was 3.0 keV, the beam current $\sim$100 pA. At a pixel dwell time of 6 μs and a pixel size of 16.5 nm $\times$ 16.5 nm this resulted in an electron dose of about 14 electrons nm^{-2}, not accounting for skirting due to low-vacuum operation. The chamber was kept at a pressure of 130 Pa hydrogen to prevent charging. The electron microscope was equipped with a custom-made microtome[24], which allows the repeated removal of the block surface at a cutting thickness of $\geqq$25 nm. A total of 3,200 consecutive slices were imaged, leading to a data volume of 8,192 $\times$ 7,072 $\times$ 3,200 voxels (a 4 $\times$ 4 mosaic of images 2,048 $\times$ 1,768 pixels in size). As the edges of neighbouring mosaic images overlapped by $\sim$1 μm, this corresponds to a physical size of about 132 μm $\times$ 114 μm for each slice and a total thickness of 80 μm. Note that stitching led to substantial shear (about 4 degrees) in z.

The cutting speed was 0.5 mm s^{-1}. To avoid chatter and ensure even cutting, the diamond knife (facet angle 50°, clearance angle 20°, Diatome) was vibrated along the knife-edge direction with a frequency of $\sim$12 kHz using a small piezo actuator integrated into the knife holder[50]. Focus and astigmatism were continually monitored (using the 'heuristic algorithm' described previously[51]) on the basis of acquired images and automatically adjusted. After each cut, a low-resolution overview image was acquired and used to automatically detect cutting debris on the surface. If debris was detected, the knife was passed over the surface with 40-nm clearance in an attempt to remove the debris. Consecutive slices were aligned offline to sub-pixel precision by Fourier shift-based interpolation, using cross-correlation-derived shift vectors. Note that the sub-volume inside the data set that contains valid data is a rhomboid.

Skeletonization. The data set was prepared as described previously[25] for crowd-sourced skeletonization by trained human annotators, which were specifically recruited from the local student population. This is different from some other 'citizen science' projects but encountered similar problems, such as the need to establish a mechanism for cross-validation. The data were visualized and annotations were captured using the KNOSSOS program[25] (http://www.knossostool.org). First, all somata in the inner plexiform and ganglion cell layers were identified and classified as ganglion, amacrine, bipolar, horizontal and glia cells, using the location of each soma and the types of neurites emerging from it. Then, starting from the soma, each neuron was traced, by multiple tracers (6, 4 and 4 for ganglion cells, bipolar cells and amacrine cells, respectively). Tracings were then consolidated using RESCOP[25] with the following refinements: all edges within 3 μm of the soma centre were eliminated, no edges were eliminated between 3 and 10 μm, and, except for somata in the ganglion cell layer, branches were allowed to pass 15 μm only if their multiplicity (pro votes) compared to the maximum multiplicity of any branch leaving the same soma (total votes) was acceptable according to the voting rules[25]. Type-grouped skeletons were visually scanned using Amira (VSG, Merignac Cedec) and KNOSSOS. For 34 apparently aberrant branches, their originating branch points were inspected in the raw data, and removed if erroneous (12 cases). Density profiles were calculated by collecting the edge centres into 50-nm-wide bins using the length divided by the total vote count[25] for each edge as the weight. Histograms were normalized and used to calculate the quartiles.

Type classification. Cells were visually inspected using views as shown in Fig. 1b. The morphological criteria used were the neurite density with depth in the IPL and the lateral branching pattern. Connectivity information was used to subdivide CBC5s (see below). If a cell could not be grouped with at least one other cell it was not assigned to a type and instead added to the 'orphan' category, even when showing a discernible neurite morphology (Supplementary Data 6). The contact data for the 110 orphan cells are not shown in Fig. 4 but are included in Supplementary Data 4 and 5. We refer to the types here by their column/row index in the type matrix. Supplementary Data 4, sheet 3, and Supplementary Data 7

provide translation between, respectively, the different indices for individual cells and between type indices, type identifiers, and common type names.

The classification of the cells proceeded as follows. The neurite ramification pattern in the IPL, particularly its distribution along the light axis and its overall lateral size, was used first. We don't usually comment when cells obviously cluster into a type by those criteria (for example, types 9, 33 and 51). Unless otherwise indicated, percentage numbers represent position along the light axis. As the boundaries of the IPL (0%, 100%), we defined the points where the total skeleton density falls below 15% of its maximum. We use the point where the skeleton densities of ON and OFF bipolar cells cross over (46.5%), as the ON/OFF boundary. In some cases (types 58–62, corresponding to CBC1–4, and in one case for the CBC5A versus CBC5R distinction) we used, in addition, tiling (the lateral overlap between neurites in the plane of the retina).

First, we identified the ON and OFF SACs (types 33 and 51). We next considered all remaining cells that had their somata in the GCL. Because we were initially not sure how reliably the axon could be detected, we did not use the presence of an axon as a criterion to distinguish ganglion cells from displaced amacrine cells. In all but one of the cells classified as ganglion cells an axon was found eventually. We begin with the actual ganglion cells (types 1–12), postponing the discussion of displaced amacrine cells (types 27, 43, 51 and 56–57).

There are three clearly bistratified ganglion cell types (2, 8 and 9) that extensively ramify in both ON and OFF layers. Only type 2 has one of its bands immediately adjacent to the INL, whereas the lowest band of type 8 is well separated from the INL. Additional discrimination was provided by bands at 50% and 70% for types 2 and 8, respectively. Only one of the type-8 cells shows all aspects of the dendritic tree, whereas the other two cells are presumably missing parts of the dendrite inside the reconstructed volume but share enough features to put them into the same class. Type 9 is the ON/OFF DSGC.

Type 6 could be called bistratified but the space between the bands still contains a lot of neurite. The two bands are just inside the choline acetyltransferase (ChAT) bands, which is where the SACs (types 33 and 51) and DSGCs (type 9) ramify. Types 7 and 10 both have only one band straddling the ON/OFF border, but 7 has numerous branches going all the way to the INL. Types 7 and 10 probably correspond to the two subtypes labelled in the TYW3 mouse[34].

Next we considered cells that ramify mostly in the OFF (types 1, 3, 4 and 5) or the ON layer (types 11 and 12). Among those, only types 1 and 3 ramify all the way up to the INL (a slight dendritic resemblance to type 27, a displaced amacrine cell, can be resolved by looking at the lateral (in-plane) branching pattern, which is much more tortuous for 27). Type 1 has multiple branches emerging directly from the soma but type 3 only has a single one. Type 5 has a much denser in-plane branching pattern than type 4. Type 11 ramifies further towards the GCL than type 5 and is broader than type 10. Type 12 is the only ganglion cell ramifying in a single band adjacent to the GCL.

Among the amacrine cells we started with the narrow-field types (13–24). Types 18, 20 and 24 all reach deep into the ON layer and have bands in both the ON and OFF layers, which was used to separate them from 23 and 22, with no bands in the OFF layer, even though the variability of the OFF band in type 20 made it difficult to distinguish type 20 and 22 cells, possibly causing some misclassification. Type 18 shows a sharp band at about 70% and a broader band touching the INL. Types 13 and 14 were difficult to distinguish, but 14 has a clearer gap to the INL and a less dense dendrite. Types 16 and 17 differ in lateral size. Some overlap between 16 and 15 cannot be completely ruled out but most type 15 cells are shorter and end mostly in a dense band. Types 19 and 21 differ in lateral size (21 and 42 may be the same type).

Next we considered cells (types 25, 28, 30–32, 37, 39, 41, 47, 53 and 57) in which the branching pattern suggested wide fields, for example, because only few of their branches ended inside the sample. Many of these cells (types 25, 37, 39, 41, 47 and 53) show a sharp lamination in depth. Only type 25 ramifies close to the INL. Type 30 is more strongly branched than 28 and ramifies broadly in depth, unusual for wide-field cells. Type 28, unlike type 30, has two branches leaving in opposite directions. Type 31 dendrites, uniquely among the cells reconstructed, go off into a narrow segment. Type 32 ramifies in the OFF ChAT band, but branches differently from the OFF SAC (type 33). Type 39 has only a single primary branch, whereas type 41, which stratifies at almost the same depth, has several. Note that types 37, 41/39 and 47 subdivide the space between the ChAT bands into three equal sublaminae.

The remaining amacrine cells are medium-field cells (types 26, 27, 29, 33–36, 38, 40, 42–46, 48–52 and 54–57), including the unmistakable SACs (types 33 and 51, see above). Types 34 (an interplexiform cell), 49 and 52 uniquely reach all the way across the IPL. Type 49 has the very distinctive 'waterfall' anatomy and type 52 lacks the sharp band right outside the INL of type 34. Types 35 and 38 were distinguished by how far their dendrites reached towards the GCL. Types 48 and

50 differ in primary dendrite shape and in-plane size but may still be the same type. Type 45 has more primary dendrites than type 54.

To classify bipolar cells (types 58–71) we first tried to establish similarity to the types described previously[28]. The correspondence was mostly obvious for RBCs (type 71) and ON CBCs (types 63-70)—see the main text for CBC5 (types 63–65) and XBC (type 66)—but rather difficult for OFF CBCs (types 58–62).

First, all OFF bipolar cells were sorted using the seventy-fifth percentile of the cumulative skeleton density in depth (starting from 0%) then, the lower 58.2% (their prevalence; see Table 1 in ref. 28) of cells were placed in the CBC3A/3B/4 and the remainder into the CBC1/2 category. The former was then sorted by the twenty-fifth percentile. Because this distribution was not clearly separable (consistent with the CBC4 width being smaller and more variable than drawn previously[28]), we began to collect the CBC4 cells starting at the highest twenty-fifth percentile numbers, adding cells consistent with the mosaic until the required prevalence was reached. The same procedure, now using the axonal coverage area, was used to separate CBC3A from B, reported to be larger CBC3A (ref. 28) and CBC1/2 using the spread in depth of the axon (twentieth to eightieth percentile). Finally, all mosaics were inspected again, six cells were reassigned, and one cell (cell 927, Supplementary Data 6) was moved to the 'orphan' group as it did not fit into any of the mosaics. In the resulting grouping, types 60–62 show a ramification-free zone adjacent to the INL that is lacking in types 58 and 59. Type 59 dendrites, if anything, are closer to the INL than type 58 dendrites. Type 62 ramifies slightly more widely in depth than types 60 and 61. Type 61 tends to be smaller than type 60.

Segmentation. A feed-forward convolutional neural network[27] was trained to classify connectedness (roughly a probability) between voxels sharing a face (the Matlab code and the network weights are in Supplementary Data 8). Several sub-volumes (each $100 \times 100 \times 100$ voxels in size) were fully segmented using KLEE and served as the initial training data, which was gradually augmented by semi-automatically segmented volumes (proofread segmentations generated with earlier network versions), yielding a final training set of 12 substantial volumes ranging from $(128 \text{ voxels})^3$ to $(240 \text{ voxels})^3$ (more than 800 million example image patches, including translations and rotations). The network contained 5 hidden layers with 10 feature maps each and was trained for over 5.5 million mini-batch gradient update steps until convergence, corresponding to many central processing unit (CPU) months, in a greedy and supervised layer-wise manner using a modified version of MALIS[49], modified to assign equal weight to each segment (S.C.T. *et al.*, manuscript in preparation). All filters were $7 \times 7 \times 7$ voxels in size and used a logistic sigmoid nonlinearity. After classifying voxel connectedness for the whole data set, segmentation was as follows. First, the voxels were clustered using a threshold of 0.9999. Clusters with ≥ 10 voxels were used as seeds and grown to threshold of 0.999. Unconsumed voxels were clustered using the same threshold, followed by seed selection and growth, now to 0.99. This procedure was repeated using thresholds of 0.98, 0.96, 0.94, 0.92, 0.9, 0.85, 0.8, 0.7, 0.6, 0.5, 0.4 and 0.2, and resulted in the assignment of each voxels in the data set to a supervoxel (on average 517 voxels), which were now merged using the following criteria: first, objects larger than 36 voxels were merged with each other if the boundary classifier averaged across their interface was above a threshold that was gradually lowered from 0.95 to 0.75 in linear steps of 0.05. In the next phase, only objects of unequal size were allowed to merge. The 'forbidden' size intervals (in voxels) and the interface thresholds for each step were: 2,000–200, 0.65; 2,000–200, 0.6; 2,000–200, 0.55; 2,000–200, 0.55; 2,000–400, 0.6; 2,000–800, 0.6; 2,000–1,600, 0.6; 2,000–1,600, 0.6; 2,000–1,600, 0.6; 5,000–2,000, 0.6; 10,000–3,000, 0.6; 20,000–4,000, 0.6; 25,000–5,000, 0.6; 30,000–6,000, 0.6.

This increased the average segment size to 2,443 voxels. Segments were then assigned to that skeleton that had the most nodes in the segment (only a small fraction contained nodes from more than one skeleton). All segments assigned to a skeleton comprise the volume reconstruction of the corresponding cell. The volume fraction erroneously assigned was estimated by summing the volume of all segments that contained multiple skeletons, weighted by the fraction of minority nodes in the segment and divided by the total volume of segments assigned to any skeleton.

Contact detection. To quantify contacts between segments, segment-to-segment overlap matrices were calculated between the original segmentation and versions shifted by one voxel, respectively, in the *x*, *y* and *z* directions. The resulting three collections of overlapping voxels were combined and classified and grouped into 'contacts' (Supplementary Data 8) using a dilation-based proximity measure. The contact areas were calculated using the following weights (nm^2) depending on according to the combination of direction sets they occurred in: 412.5 (*x* or *y*), 272.25 (*z*), 583.3631 (*x* and *y*), 494.2432 ((*x* or *y*) and *z*), 643.7644 (*x*, *y* and *z*). This corrects for the anisotropy in voxel size and to some extent for the error introduced by the angle of the contact surface. For surfaces perpendicular to one of the principal axes, the face diagonals, or the space diagonal this estimate is exact.

Error estimation. To probe the frequency of missed contacts (false negatives) we selected 100 random locations on one skeleton (cell 17, gc36-51, W3) and searched for true contacts with an, according to the cell–cell matrix, highly connected cell (cell 344, ac34-84). All 16 true contacts found were also found by the automated detection routine. To estimate the false positive rate we randomly selected 20 of the 7,217 contacts that the same ganglion cell made with other cells and visually inspected the corresponding locations in the raw data. In one case no actual contact existed (a piece of debris was erroneously attributed by the segmentation routine to cell 344).

Sizes for synaptic and non-synaptic contacts. Synaptic and non-synaptic contacts in the conventionally stained data set (k563) were selected and their contact area determined in one of two ways. (1) Starting from a bipolar cell axon terminal, a synaptic ribbon was located (Fig. 3d), the two postsynaptic dyadic partners were found and their class determined, using the presence or absence of synaptic vesicles (found in amacrine but not ganglion cell dendrites). All three dyadic partners and, in addition, a nearby non-synaptic contact were manually reconstructed using the KLEE software tool (M.H. *et al.*, manuscript in preparation) in a region including all three contacts. The contact areas were determined as follows. Surface triangulations were generated for each volume reconstruction, then for each triangle it was determined whether there was another object within 144 nm above it, next the contact area with this object was calculated as the sum over all hits in that object weighted by the triangle areas. (2) All contacts with bipolar cells were reconstructed on several pieces of ganglion cell dendrite, quantified, and classified as synaptic when a ribbon was present and non-synaptic otherwise. Classification, segmentation and contact detection were performed independently for each member of a set of overlapping cubes (257 voxels on a side), one cube for each interior data cube (128 voxels on a side). Each of those cubes overlaps one data cube completely and 26 cubes partially. To avoid double counting, we counted a contact only when the largest part of the contact was inside the completely overlapped (central) data cube.

50. Studer, D. & Gnaegi, H. Minimal compression of ultrathin sections with use of an oscillating diamond knife. *J. Microsc.* **197**, 94–100 (2000).

51. Binding, J., Mikula, S. & Denk, W. Low-dosage maximum-*a-posteriori* focusing and stigmation. *Microsc. Microanal.* **19**, 38–55 (2013).

ARTICLE

A visual motion detection circuit suggested by *Drosophila* connectomics

Shin-ya Takemura[1], Arjun Bharioke[1], Zhiyuan Lu[1,2], Aljoscha Nern[1], Shiv Vitaladevuni[1], Patricia K. Rivlin[1], William T. Katz[1], Donald J. Olbris[1], Stephen M. Plaza[1], Philip Winston[1], Ting Zhao[1], Jane Anne Horne[2], Richard D. Fetter[1], Satoko Takemura[1], Katerina Blazek[1], Lei-Ann Chang[1], Omotara Ogundeyi[1], Mathew A. Saunders[1], Victor Shapiro[1], Christopher Sigmund[1], Gerald M. Rubin[1], Louis K. Scheffer[1], Ian A. Meinertzhagen[1,2] & Dmitri B. Chklovskii[1]

Animal behaviour arises from computations in neuronal circuits, but our understanding of these computations has been frustrated by the lack of detailed synaptic connection maps, or connectomes. For example, despite intensive investigations over half a century, the neuronal implementation of local motion detection in the insect visual system remains elusive. Here we develop a semi-automated pipeline using electron microscopy to reconstruct a connectome, containing 379 neurons and 8,637 chemical synaptic contacts, within the *Drosophila* optic medulla. By matching reconstructed neurons to examples from light microscopy, we assigned neurons to cell types and assembled a connectome of the repeating module of the medulla. Within this module, we identified cell types constituting a motion detection circuit, and showed that the connections onto individual motion-sensitive neurons in this circuit were consistent with their direction selectivity. Our results identify cellular targets for future functional investigations, and demonstrate that connectomes can provide key insights into neuronal computations.

Vision in insects has been subject to intense behavioural[1], physiological[2] and anatomical[3] investigations, yet our understanding of its underlying neural computations is still far from complete. One such computation, ethologically highly relevant, is motion detection, which is thought to rely on the comparison between signals offset in space and time[4–6] (Fig. 1a, b). Yet, despite being the focus of theoretical and experimental investigations for more than half a century[7], the exact mechanism underlying this computation remains a mystery. A central impediment towards unravelling this mechanism has been our incomplete knowledge of the relevant neurons and synapses.

In the fly, visual processing begins in the optic lobe, composed of four retinotopically organized neuropils. Each is an array of repeating modules corresponding to the hexagonal lattice of ommatidia in the compound eye (Fig. 1c). Each module of the first neuropil, the lamina, contains a repeating circuit[8,9] receiving inputs from six photoreceptors detecting light from the same location in the visual field. The output cells of each lamina module project to a corresponding module of the second neuropil, the medulla (Fig. 1c). Each medulla module, called a column (Fig. 1d), is also thought to contain stereotypic circuits[10]. These columns, in turn, innervate two downstream neuropils, the lobula and lobula plate (Fig. 1c).

Responses to local motion must be computed at least partly within the stereotypical circuits of the medulla columns. Indeed, the medulla is the first neuropil with movement-specific activity[11], and, directly downstream of the medulla, lobula plate tangential cells (LPTCs) integrate local motion signals to produce wide-field motion response[12,13]. However, up until now, the lack of a medulla connectome has frustrated investigations of local motion detection.

Semi-automated connectome reconstruction

To provide a reliable foundation for computational modelling and identify targets for electro/optophysiological recordings, we attempted a complete, dense reconstruction of the chemical synaptic connectivity within the medulla using electron microscopy, the gold standard of neuroanatomy[14]. Given the time-consuming nature of such reconstructions, we wanted to determine the smallest medulla volume, reconstruction of which would allow us to identify a circuit underlying the computation of local motion. Both directional turning responses[15] and electrophysiological responses in LPTCs[16] can be elicited in flies by sequential stimulation of two photoreceptors corresponding to adjacent points anywhere in the visual field[17]. This suggests that some repeating component of the motion detecting circuit must be present within any two adjacent medulla columns. We therefore decided to reconstruct all the synaptic connections among neurons within a single reference column, as well as all the connections between the reference column and neurons within six nearest-neighbour columns (Fig. 1d).

Because manual reconstruction of even a seven-column volume would be prohibitively time-consuming[18], we developed a semi-automated reconstruction pipeline[19] and applied it to the medulla volume (Fig. 2, Methods and Supplementary Data 1), reconstructing 379 cells (Supplementary Fig. 1 and Supplementary Video 1).

To map our reconstruction onto the existing body of knowledge, we assigned these cells to previously proposed cell types[20] by comparing the shapes of reconstructed arbors (Supplementary Fig. 1) with those reported from light microscopy using Golgi impregnation or genetic single-cell labelling (Fig. 2e, f and Supplementary Methods). Because there were several reconstructed examples for almost all neuronal types (Supplementary Fig. 1 and Supplementary Table 2), it was possible to characterize the common structural features of each type. In many cases, this allowed us to match unequivocally a reconstructed cell with a Golgi impregnate[20], for which it was then named (Supplementary Methods). However, there was also a subset of cell types for which a Golgi counterpart could not be found but which we validated using isomorphs from genetic single-cell labelling. We named these cell types Mi13, Mi14, Mi15, TmY14, Dm9, and Dm10 (Supplementary Fig. 2). In total, from the collection of 379 reconstructed cells (Supplementary Fig. 1) we were able to classify 290 of them into 56 cell types (Supplementary Table 2).

[1]Janelia Farm Research Campus, HHMI, Ashburn, Virginia 20147, USA. [2]Department of Psychology and Neuroscience, Dalhousie University, Halifax, Nova Scotia B3H 4R2, Canada.
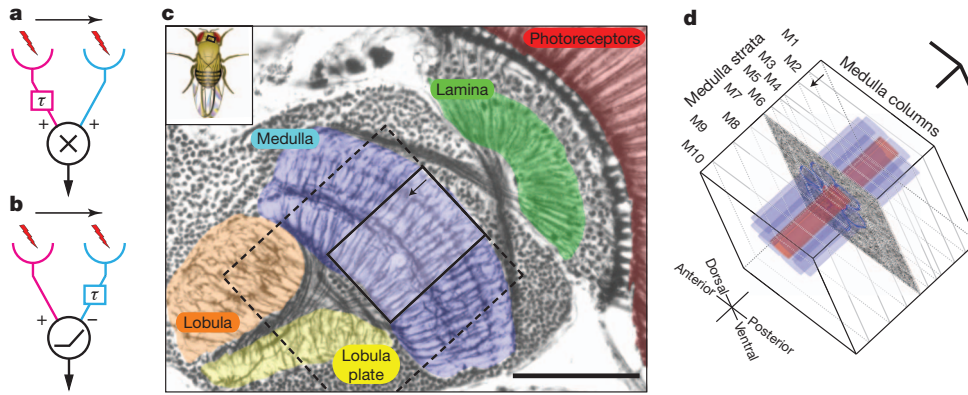
Figure 1 | Motion detection and the *Drosophila* visual system. a, Rightward motion component of the Hassenstein–Reichardt EMD[4] model. Light input (lightning bolt) into the left channel (magenta) is transmitted with an additional delay, τ, relative to that into the right channel (cyan). For a rightwards-moving object, signals from both channels will arrive at the multiplication unit closer in time to each other, and therefore become nonlinearly enhanced (and vice versa for leftwards-moving objects). As a result, the model responds preferentially to rightward motion. **b,** Alternative Barlow–Levick-like EMD[6] model, also preferring rightward motion. Note that the inputs are combined with opposing signs and the delay is now in the right (cyan) channel. **c,** Bodian silver-stained horizontal section[45] of the *Drosophila melanogaster* visual system revealing the four neuropils of the optic lobe. The medulla region of interest (solid rectangle, expanded in **d**) and the wider imaged volume (dashed rectangle) used to trace into the lobula plate are shown schematically. **d,** The 37 μm × 37 μm medulla region of interest is centred on the reference column (red) and six surrounding nearest-neighbour columns (blue). The medulla has ten strata (M1–M10) defined by the arborizations of its cell types. Scale bars, 50 μm (**c**) and 10 μm (**d**, in all three directions).
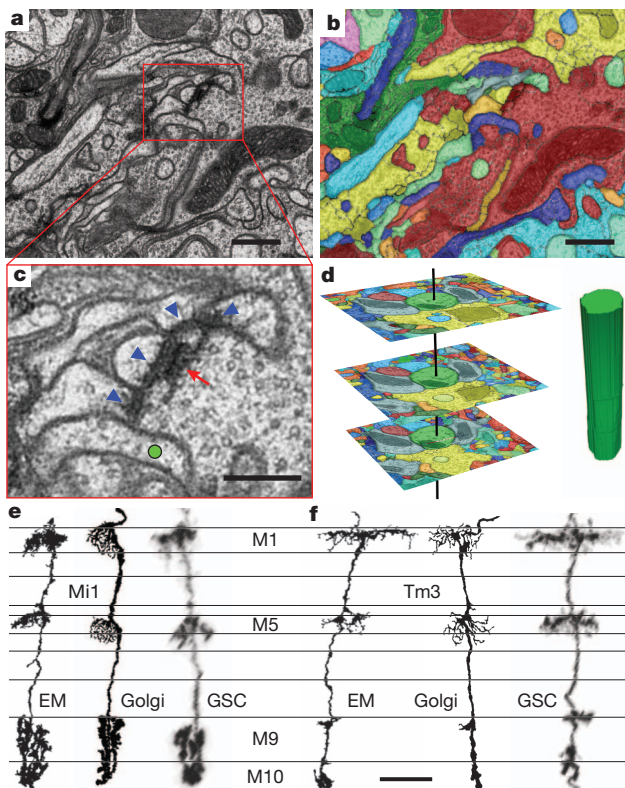
Figure 2 | Connectome reconstruction using serial-section electron microscopy. a, A representative micrograph, one of 2,769 from the electron microscopy series. **b,** Proofread segmentation of the micrograph in **a** into neurite profiles (single colours). **c,** Synapses comprise a presynaptic process containing a T-bar ribbon (red arrow) and associated neurites with postsynaptic densities (PSDs) (blue arrowheads) adjacent to the T-bar. A non-synaptic process (green circle) lacks a PSD (in both this and other section planes containing this T-bar). **d,** Neurites are reconstructed by linking profiles in consecutive sections (left), to construct a 3D object (right). **e,** An example of a neuron reconstructed from electron microscopy (left), identified by comparison with the Golgi impregnated cell (centre)[20] as type Mi1 and cross-validated by a corresponding genetic single-cell (GSC)-labelled neuron (right) (Supplementary Methods). **f,** Same as **e** for cell type Tm3. Scale bars, 500 nm (**a**, **b**), 250 nm (**c**) and 10 μm (**e**, **f**).

To reveal the connections between the 379 reconstructed neurons, we identified pre- and postsynaptic sites and assigned them to their respective parent cells. Within the reference column and its immediate surround, we annotated 10,093 presynaptic sites and 38,465 associated postsynaptic sites (3.8 ± 1.2 (mean ± s.d.) per presynaptic site) (Fig. 2c and Supplementary Table 1). Although presynaptic T-bars typically fell onto proofread profiles of neurons, postsynaptic sites usually fell onto isolated profiles, unassigned to any neuron. Thus, it was necessary to trace the dendrite containing each post-synaptic site back to a parent cell. This postsynaptic tracing was extremely challenging as *Drosophila* neuron dendrites branch elaborately and, indeed, can be thinner than the section thickness.

The challenging postsynaptic tracing led to (1) some erroneously identified synaptic contacts, and (2) a high fraction (~50%) of contacts that could not be traced to their parent neuron and were therefore unidentified. To increase our confidence in the identified contacts (1), we had two proofreaders trace every postsynaptic site (Methods), and only accepted into the connectome those contacts that both proofreaders identified independently. By contrast, it was not possible to reduce the number of unidentified contacts experimentally (2). However, we were still able to construct a connectome valuable for inferring function because we found that, within the medulla, connections of high weight (that is, high number of synaptic contacts per connection) both capture a large fraction of the total connection weight and can be identified with high fidelity. Indeed, the distribution of connection weight in our connectome is heavy tailed (Fig. 3b, inset, and Supplementary Fig. 3), as has been found in other organisms[21,22]. Also, assuming that synapses are equally difficult to proofread, we found that any strong connection (with >5 synaptic contacts) will be identified with >95% probability. Therefore, in the resulting connectome, 8,637 synaptic contacts are precisely identified, and all strong connections are represented.

The connectome module and its pathways

To identify pathways performing local computations such as motion detection, it was necessary to generate a more convenient abstraction of the full connectome. Because we expect that the circuits of interest repeat within each column, we extracted from the medulla connectome a periodic module of connections between identified cell types that arborize in every column. These include both so-called synperiodic cell types with single neurons in every column of the medulla[23], and cell types with several members within each column, which we
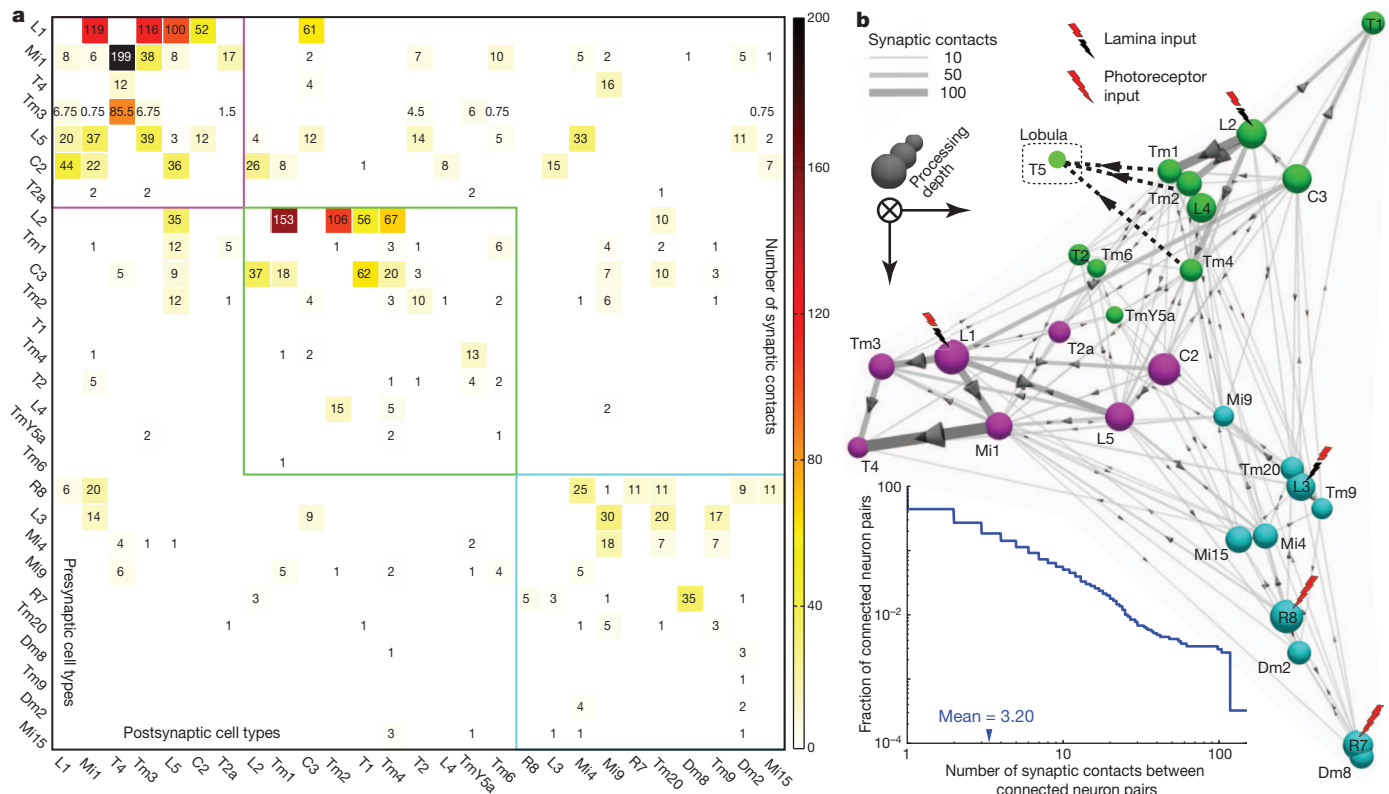
Figure 3 | Medulla connectome module. a, Synaptic connectivity matrix for modular cell types assembled from 2,495 synapses (Supplementary Table 1). Three pathways, identified via the Louvain clustering analysis[46], are labelled by coloured boxes. They are named by their primary input neuron(s): the L1 (magenta), L2 (green) and L3/R7/R8 (cyan) pathways. The pathways are ordered by the total number of connections within a pathway, in descending order, and the cell types, within each pathway, are ordered by the sum of their pre- and postsynaptic connections to and from other cell types within their pathway, also in descending order. **b**, Medulla connectome module as a 3D graph. Cell types with stronger connections are positioned closer to each other, using the visualization of similarities layout algorithm[47]. Three spatially segregated groups are observed that closely match the pathways identified through clustering (colouring of spheres). The dominant direction of signal flow is oriented into the page[22]. Inset in **b** shows the fraction of synaptic connections within the full connectome having a connection weight greater than indicated.

term ultraperiodic. We do not include infraperiodic tangential or local amacrine-like cells even if they have arborizations in every column because this cannot be determined unambiguously from our electron microscopy reconstruction (Supplementary Table 2). We used the existence of multiple representatives from adjacent columns within our electron microscopy reconstruction (Supplementary Fig. 1) to identify 25 cell types as synperiodic, as well as two cell types, Tm3 and T4, as ultraperiodic (with 1.5 and 4 cells per column, respectively) (Supplementary Table 2). We termed these 27 cell types modular.

Assuming that connections between modular cells are stereotypical between columns, we constructed the repeating circuit module by finding all connections between these cell types (Methods). Unlike sparse reconstructions, the resulting connectome module (Fig. 3a) accurately captures not only the presence but also the absence of strong connections between any two cell types.

To determine which neurons could be involved in different local computations, we dissected the connectome module into three separate signal processing pathways, using both a clustering and a layout algorithm (Fig. 3). We recognized them as the previously identified pathways, those of L1, L2 and L3/R7/R8. The downstream targets of R7 and R8 have previously been implicated in colour vision. Because colour pathways are separated into differing columns receiving inputs from either pale or yellow ommatidia[24], we expect that they should rely on infraperiodic cell types omitted from our connectome module. Therefore, the fine structure of the L3/R7/R8 pathway will be revisited elsewhere.

The remaining L1 and L2 pathways signal visual contrast, and are implicated in motion detection[7,25–29]. Behavioural experiments and

electrophysiological recordings confirm this role for L1 and L2: not only is each necessary for aspects of motion detection[29–31], but, among the cells postsynaptic to the photoreceptors R1–R6, both are also mostly sufficient[30,31] for the computation. However, L1 and L2 themselves lack directionally selective responses[7,25]. Therefore, to search for motion detection circuit(s) within the connectome module, we examined the neurons downstream of L1 and L2 in more detail.

Candidate motion detection circuit

Several lines of evidence indicate that motion information computed downstream of L1 and L2 is relayed to the lobula plate via cell types T4 and T5 (ref. 25). First, recordings from LPTCs in fruitflies with genetically silenced T4 and T5 demonstrate that at least one of these columnar cell types is necessary to detect direction selectivity[32]. Second, T4 and T5 cells in *Drosophila* each comprise four subtypes differentiated by the lobula plate layer in which their axons arborize[20]. Third, each of these four layers within the lobula plate exhibits activity in response to wide-field stimuli moving in a particular direction: downwards, upwards, backwards and forwards (Fig. 4b, e), revealed by their uptake of deoxyglucose[11,27]. Finally, dendrites of LPTCs with different motion preference co-occupy the lobula plate layers corresponding to their directional preference[20] and, in addition, receive direct synaptic connections from T4 terminals[28]. Collectively, these data suggest that each subtype of T4/T5 forms the output of motion detection circuits signalling a particular direction of motion.

Next, we argue that the direction-selective outputs of T5 and especially T4 are computed largely independently of each other. Consistent with stratum-overlap analysis in *Drosophila*[26] and large flies[25], our
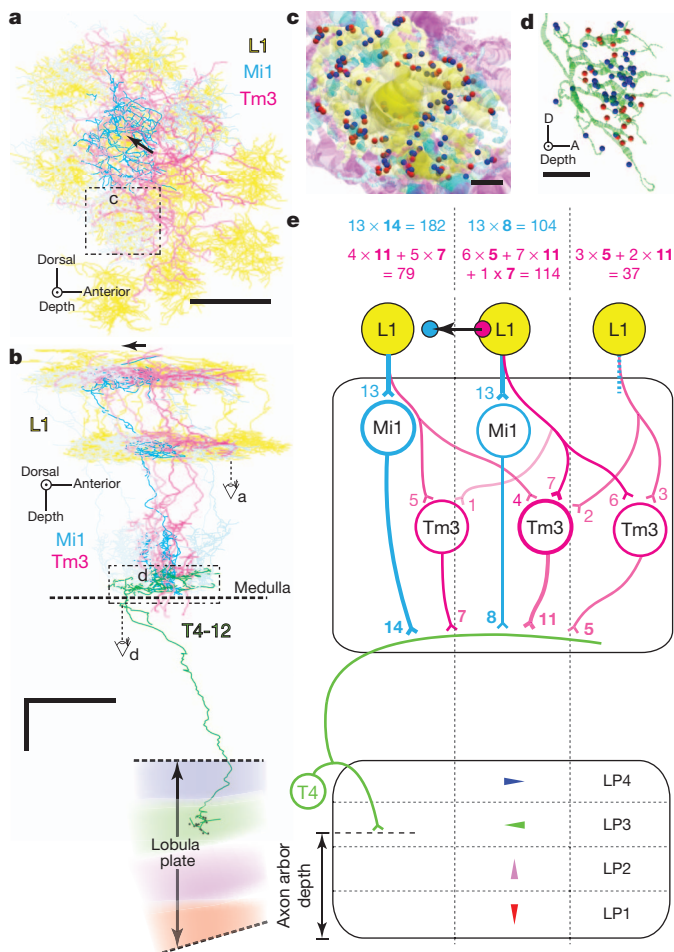
Figure 4 | Spatial displacement of Mi1- and Tm3-mediated inputs onto a single T4 cell (T4-12). **a**, Bottom view of dendrites of the Mi1 (cyan) and Tm3 (magenta) neurons presynaptic to T4-12, overlaid on the array of L1 axonal terminals (yellow). The colour saturation for each dendritic arbor reflects the number of synaptic contacts made onto T4-12 (see **b** and **d**). Arrow shows the displacement from the Tm3 centre of mass to the Mi1 centre of mass computed as illustrated in **e**. **b**, Side view of T4-12 and its presynaptic Mi1 and Tm3 cells. Direction preference for a T4 cell (coloured to match the directional preferences in **e**) is determined by the lobula plate arborization layer of the axon terminals (T-bars in black). **c**, Enlargement (dashed rectangle in **a**) showing reconstructed neurites of Mi1, Tm3 and L1 cells (without the weighted colours in **a**), and their synaptic contacts (L1 → Mi1: blue; L1 → Tm3: red). **d**, Reconstructed dendritic arbor of T4-12 with synapses from Mi1 (blue) and Tm3 (red). **e**, Cartoon of inputs to a single T4 cell through Mi1 and Tm3 cells. Mock synaptic weights illustrate how the receptive fields were computed. The centre of mass of Mi1 (or Tm3) component, blue (or red) circle, is computed by placing the mass corresponding to the compound synaptic weight from L1 through Mi1 (or Tm3) to T4 at the centre of the corresponding column. Scale bars, 8 μm (**a**), 8 μm (**b**), 1 μm (**c**) and 4 μm (**d**).

connectome (Fig. 3a, b) indicates that T4 belongs to the L1 pathway. Conversely, although we did not trace into the lobula, the dendrites of T5 cells co-occupy layers in the lobula with the axon terminals of neurons such as Tm1 (refs 20, 25), Tm2 (ref. 20) and Tm4 (ref. 20), which belong to the L2 pathway (Fig. 3b). Electrophysiological[30] and behavioural[29] evidence indicate that the L1 and L2 pathways in the connectome module (Fig. 3a, b) are computationally independent and correspond to ON and OFF pathways in the visual systems of vertebrates[30]. Furthermore, most of the connections from the L2 to the L1 pathway arrive via L5 (Fig. 3a), a cell type not implicated in motion detection[33]. Hence, we decided to search for a motion detection circuit downstream of L1 converging on T4 cells. This decision was made despite electrophysiological evidence showing a lack of direction

selectivity in T4 cells[34], but since the completion of our work, has been supported by calcium imaging of T4 cells[35].

To identify candidate elements of a motion detection circuit bridging between L1 and T4, we took advantage of the fact that motion detection is both fast and robust to noise[16,17], and consequently should be implemented by both feedforward and strong connections. Our dense electron microscopy reconstruction identified five cell types with significant input from L1: Mi1, Tm3, L5, C2 and C3 (Fig. 3a). Cell types Mi1 and Tm3 are the two largest recipients of L1 input, together accounting for more than half of the synaptic contacts of L1. In turn, Mi1 and Tm3 together contribute >80% of presynaptic inputs to T4 (including all inputs from both modular and non-modular cell types), thus forming the two strongest paths from L1 to T4. By contrast, cell type C3 contacts T4 with an order of magnitude fewer synapses than Mi1 and Tm3, suggesting that its contribution is far weaker. Finally, cell types L5 and C2 have no synapses directly with T4. These features lead us to suggest that Mi1 and Tm3 are the primary substrates for robust and fast motion detection within the L1 pathway.

Anatomical receptive fields of T4 cells

To explore further whether Mi1 and Tm3 cells converging on T4 cells could constitute the two arms of a correlation-based motion detector (Fig. 1a, b), we examined whether the motion axis defined by these inputs onto a particular T4 is consistent with its preferred direction, as measured by its outputs. This output-preferred direction was determined for 16 T4 neurons by tracing their axons into the lobula plate and identifying their arborization layer[11,27] (Fig. 4b, e).

To compare the preference of T4 output with the motion axis arising from its inputs, we constructed the input motion axis by analysing the numbers of contacts from individual Mi1 and Tm3 cells onto the T4 neuron in question. We found that each T4 receives inputs from several Mi1 and Tm3 neurons suggesting that, unlike the circuits in Fig. 1a and b, several points in the visual field provide inputs into each arm of the motion detector (Fig. 4a, b, e). This observation is supported by the structure of sampling units inferred from recordings in blowfly H1 (ref. 36), which receives inputs from LPTCs[37]. We therefore needed to characterize the inputs to each T4 as Mi1- and Tm3-mediated receptive field components, mapped into the visual field. To do this, we traced synaptic connections from L1 terminals in 19 columns (the reference column and surrounding 18 columns) to the downstream Mi1 and Tm3 cells and then from the Mi1 and Tm3 neurons onto T4 cells that receive input from the reference column (which also happen to number 19). The resulting receptive fields (Fig. 4a and Supplementary Fig. 4) show the T4 inputs mapped as if on the L1 array and hence into the visual field.

For all T4 cells, the Mi1- and Tm3-mediated components of the T4 receptive field overlap substantially with one another (Supplementary Fig. 4). Indeed, the centres of mass of the two components are displaced less than one inter-ommatidial distance (Fig. 5a). However, for 15 of the T4 cells, this displacement is still significantly greater ($P < 0.05$) than would have been obtained by chance from tracing errors. Such a small displacement magnitude relative to the widths of the receptive fields agrees with previous evidence inferred from blowfly H1 recordings and has been justified theoretically[36].

Is the direction of displacement between the Tm3 and Mi1 receptive-field components for a T4 neuron consistent with the directional preference of the neuron, as defined by the depth of its terminal axonal arborization in the lobula plate? Assuming that the direction of the displacement is determined from the Tm3 to Mi1 component centres of mass, Fig. 5b (top) shows that the direction of displacement agrees with the directional preference for three of the four lobula plate layers. The discrepancy in the direction of displacement and the front-to-back motion preference of T4 cells terminating in the fourth layer (lobula plate later 4, LP4) may be caused by neglected circuits contributing specifically to the responses of these T4 neurons. For example, behavioural evidence implicates C3 neurons in front-to-back motion detection[33],
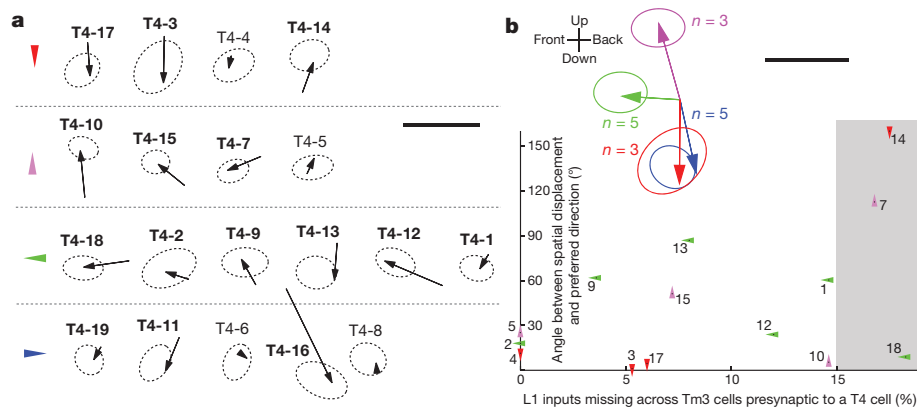
Figure 5 | Computed displacements for all T4 cells. a, Displacement vectors for each T4 neuron (*n* = 19). Neurons with significant displacement (names in bold) have 95% confidence intervals (ellipses) that exclude the origin (Methods). The vectors are in the ommatidial frame of reference (within ~30° of the visual axes). **b,** Top, mean displacement, computed from **a,** averaged over the cells with the same preferred direction of their output. Bottom, the angular difference between the spatial displacement for individual T4 neurons and the preferred direction of its output (for lobula plate layers 1–3 (LP1–3)) correlates with the fraction of missing L1 inputs (Methods). T4 cells with >15% of missing L1 inputs were omitted from the mean displacements (top). Scale bars, 0.5 (**a**) and 0.2 (**b**) of the centre-to-centre distance between adjacent facets.

and, indeed, preliminary tracing results indicate that, despite C3 cells providing an order of magnitude fewer inputs to T4 cells than Mi1 or Tm3 cells, C3 neurons target LP4 T4 cells preferentially.

Some of the discrepancy between the receptive field offset and the directional preference of individual T4 cells (Fig. 5a) innervating lobula plate layers 1–3 is probably caused by a systematic error in our reconstruction resulting from the finite size of the reconstructed region. Indeed, the discrepancy between the measured input displacement for a given T4 and its predicted direction preference correlates with the weighted fraction of missing L1 inputs onto Tm3 cells upstream of that T4 (Fig. 5b, bottom), supporting the view that the fields of peripheral T4 cells may not be fully reconstructed. In addition, some of the remaining variation in the offset orientation may also be real, given the observed 60°–90° half-width of the tuning curves obtained through calcium imaging of individual T4 subtypes[35].

Our choice to measure displacement from Tm3 to Mi1 (and not in the reverse order) seems arbitrary without including information about delays and synaptic polarity (Fig. 1a, b). To estimate a possible conduction delay, we measured both the path length and the calibre of the main axon trunks that conduct signals along the Mi1 and Tm3 cells from L1 synapses to T4 synapses and found them to be similar, within 10% of each other. Moreover, using a range of electrotonic parameters measured in other fly neurons[38], the corresponding cable delays were still only on the order of one millisecond, an order of magnitude less than that required for motion detection[29]. Furthermore, although some neurotransmitters have been identified for the cell types involved, we do not know their associated receptors and hence the resulting synaptic polarity.

In the absence of evidence for both relative delay and synapse polarity, we are free to choose the measurement direction from Tm3 to Mi1, which leads to spatial displacements consistent with the directional preference predicted by the depth of the T4 terminal in the lobula plate. Assuming that the Mi1 and Tm3 inputs to T4 are combined with the same sign, as in the Hassenstein–Reichardt elementary motion detector (EMD) model[4] (Fig. 1a), we predict that the Tm3 arm of the motion detector should introduce a longer delay than the Mi1 arm. If, however, the inputs were combined with opposing signs, as in the Barlow–Levick-like EMD model[6] (Fig. 1b), then our prediction would be the opposite. Regarding the mechanism of the delay, having the two arms of the circuit implemented by different cell types allows the possibility that the delay may be implemented biologically by means of metabotropic receptors, as reported in the vertebrate retina[39].

Exploring the reconstructed T4 cells, we identified a hitherto unrecognized feature of their medulla dendritic arbors (see, however, T4a,d in Fig. 14 of ref. 20, and N. J. Strausfeld, personal communication): the dendritic branches of each T4 neuron are oriented primarily in one direction (Fig. 6a and Methods). Moreover, the branch orientation of each T4, measured from the dendrite tips to their bases, clusters around one of four directions (Fig. 6b). These four directions, when mapped from the medulla coordinate frame onto the visual field (Fig. 6c), align with the output direction preference for each lobula plate layer (Fig. 6b). This observation allowed us to cross-validate the classification of each of the 16 T4 cells into direction preference subtypes (Fig. 6a, b and Supplementary Fig. 5). We then used this observation to infer a direction preference for the remaining three T4 cells, for which tracing into the lobula plate could not be completed (Fig. 6c).
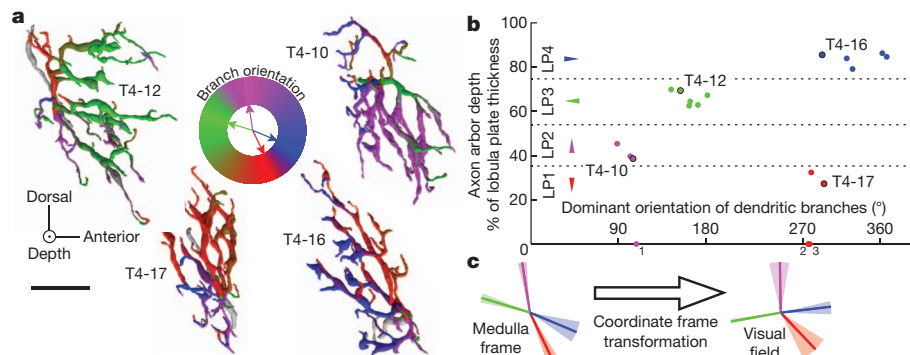

Figure 6 | Orientation of medulla dendritic arbors of T4 neurons correlates with axon terminal arborization layer in the lobula plate. a, Four representative medulla dendritic arbors of T4 cells. The colours represent local dendritic branch orientation. The colour map was constructed by assigning colours from each lobula plate layer (Fig. 4e) to the average dominant branch orientation over all neurons in each layer (arrows within colour map) and smoothly interpolating. **b,** Depth of axonal arbor of a T4 neuron within the lobula plate correlates with dominant dendritic branch orientation in the medulla (Methods). In the depth axis, four layers are labelled and the neurons within each layer are coloured as in Fig. 4e. The dominant orientations of neurons with axons not traced to the lobula plate are plotted on the *x* axis (1: T4-5, 2: T4-4, 3: T4-14), and they are coloured with the colour of the cluster to which they most likely belong (Supplementary Fig. 5). **c,** Transforming the dominant dendritic orientation (±s.e.m.) from the space defined by the array of medulla columns (in layer M10) to the directions in visual space. Scale bar, 5 μm.

Discussion

This report has introduced a new high-throughput, semi-automated pipeline for electron microscopy reconstruction, and applied it to reconstruct a connectome module comprehensively within the medulla, a neuropil that has long resisted such attempts. Furthermore, using connectomics, we identified Mi1 and Tm3 inputs to T4 neurons as the two arms of a candidate correlation-based motion detector. Although anatomy alone does not allow us to probe the nonlinear operation or the time delay, and hence distinguish between different correlation-based models[5,6,40], we were able to predict which cell type should introduce a longer delay, given their synaptic polarities (Fig. 1a, b).

Analogous to our proposed circuit downstream of L1, the connections within our electron microscopy reconstruction allow us to suggest candidate cell types—Tm1, Tm2 and Tm4—that may constitute the motion detection circuit downstream of L2. Confirmation of this suggestion must of course await dense reconstruction of the connections onto T5 neurons in the lobula.

This report has several interesting parallels with results from vertebrate retinae. First, the existence of the four subtypes of T4 cells responding to the four cardinal directions of motion is reminiscent of the four subtypes of ON–OFF directionally selective ganglion cells (DSGCs) in the rabbit retina[41]. Second, our finding that directional selectivity of T4 cells is aligned with their dendritic orientation is reminiscent of JAM-B ganglion cells[42] and starburst amacrine cells (SACs)[43] in the mouse retina. However, unlike JAM-B and SAC cells, the preferred direction of T4 cells is away from the tip of the dendrites and, unlike SAC cells but like JAM-B cells, all dendrites in one T4 point in the same direction. Third, the highly specific connections between SACs and DSGCs responsible for the directional selectivity of the latter were also demonstrated previously using large-scale connectomics[44]. However, unlike the SAC-to-DSGC circuit, the circuit we report may compute directionally selective responses from non-directional inputs.

Our identification of the candidate motion detection pathway downstream of L1 was greatly aided by the comprehensiveness of our electron microscopy reconstruction. Relative to connections estimated by arbor overlap[26], having the precise synaptic counts allows us to unequivocally establish connections (Supplementary Fig. 6). In this way, we identified Tm3 as a primary component of the motion detection circuit, a fact that escaped previous researchers owing to its minimal arborization in M10. Furthermore, relative to sparse reconstructions in other systems, for example, synaptic connections between SACs and DSGCs[44], the comprehensiveness enables us to argue both the absence of alternative pathways and the numerical importance of the proposed pathway.

The significance of the dense medulla connectome also goes far beyond the local motion detector, applying to many other visual computations. Although much remains to be done, especially to incorporate tangential and infraperiodic cells fully, our connectome does contain the columnar neurons found in every column and hence removes a long-standing block to understanding insect vision. More generally, our results illustrate that, combined with a rich collection of experimental and theoretical results, connectomes can, by identifying underlying circuits, provide key insight into neuronal computation.

METHODS SUMMARY

A fly's brain was prepared for electron microscopy by high-pressure freezing/freeze substitution, followed by embedding, and sectioning. Sections were imaged using transmission electron microscopy. Electron micrographs were assembled into a three-dimensional stack correctly representing the relative locations of the imaged structures by means of nonlinear registration. The aligned stack of grayscale images was partitioned by machine learning algorithms into subsets of pixels, each belonging to only a single neuron (see https://bitbucket.org/shivnaga/sstem). This over-segmented partition was manually inspected and agglomerated into individual neurons ('proofreading'). Presynaptic terminals and postsynaptic densities were then annotated by hand, and assigned to neurons (Supplementary Data 1). These manual steps required ~14,400 person-hours (including ~2,500 expert hours). The types of reconstructed neuron were determined by matching their shapes to examples from light microscopy. The full connectome was condensed

into a connectome module, which was partitioned into pathways using graph layout and clustering algorithms. By counting the numbers of synaptic contacts between L1, Mi1, Tm3 and T4 cells, we computed the relative displacement of the input components. We cross-validated this displacement against the direction preference of T4 cells found by tracing their axons into the lobula plate.

Full Methods and any associated references are available in the online version of the paper.

1. Heisenberg, M. & Wolf, R. Vision in Drosophila. Genetics of Microbehaviour (Springer Verlag, 1984).
2. Laughlin, S. B. Matching coding, circuits, cells, and molecules to signals: General principles of retinal design in the fly's eye. Prog. Retin. Eye Res. **13**, 165–196 (1994).
3. Strausfeld, N. J. & Nässel, D. R. in Handbook of Sensory Physiology (eds Autrum, H. et al.) 1–132 (Springer-Verlag, 1981).
4. Hassenstein, B. & Reichardt, W. Systemtheoretische Analyse der Zeit-, Reihenfolgen- und Vorzeichenauswertung bei der Bewegungsperzeption des Rüsselkäfers Chlorophanus. Z. Naturforsch. B **11**, 513–524 (1956).
5. Reichardt, W. in Sensory Communication (ed. Rosenblith, W. A.) 303–317 (MIT Press, 1961).
6. Barlow, H. B. & Levick, W. R. The mechanism of directionally selective units in rabbit's retina. J. Physiol. (Lond.) **178**, 477–504 (1965).
7. Borst, A., Haag, J. & Reiff, D. F. Fly motion vision. Annu. Rev. Neurosci. **33**, 49–70 (2010).
8. Rivera-Alba, M. et al. Wiring economy and volume exclusion determine neuronal placement in the Drosophila brain. Curr. Biol. **21**, 2000–2005 (2011).
9. Meinertzhagen, I. A. & Sorra, K. E. Synaptic organization in the fly's optic lamina: few cells, many synapses and divergent microcircuits. Prog. Brain Res. **131**, 53–69 (2001).
10. Takemura, S. Y., Lu, Z. & Meinertzhagen, I. A. Synaptic circuits of the Drosophila optic lobe: the input terminals to the medulla. J. Comp. Neurol. **509**, 493–513 (2008).
11. Buchner, E., Buchner, S. & Bülthoff, I. Deoxyglucose mapping of nervous activity induced in Drosophila brain by visual movement. J. Comp. Physiol. A **155**, 471–483 (1984).
12. Krapp, H. G. & Hengstenberg, R. Estimation of self-motion by optic flow processing in single visual interneurons. Nature **384**, 463–466 (1996).
13. Joesch, M., Plett, J., Borst, A. & Reiff, D. F. Response properties of motion-sensitive visual interneurons in the lobula plate of Drosophila melanogaster. Curr. Biol. **18**, 368–374 (2008).
14. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode Caenorhabditis elegans. Phil. Trans. R. Soc. Lond. B **314**, 1–340 (1986).
15. Kirschfeld, K. in Information Processing in the Visual System of Arthropods (ed. Wehner, R.) 61–74 (Springer Verlag, 1972).
16. Riehle, A. & Franceschini, N. Motion detection in flies: parametric control over ON–OFF pathways. Exp. Brain Res. **54**, 390–394 (1984).
17. Schuling, F. H., Mastebroek, H. A. K., Bult, R. & Lenting, B. P. M. Properties of elementary movement detectors in the fly Calliphora erythrocephala. J. Comp. Physiol. A **165**, 179–192 (1989).
18. Helmstaedter, M., Briggman, K. L. & Denk, W. 3D structural imaging of the brain with photons and electrons. Curr. Opin. Neurobiol. **18**, 633–641 (2008).
19. Chklovskii, D. B., Vitaladevuni, S. & Scheffer, L. K. Semi-automated reconstruction of neural circuits using electron microscopy. Curr. Opin. Neurobiol. **20**, 667–675 (2010).
20. Fischbach, K.-F. & Dittrich, A. P. M. The optic lobe of Drosophila melanogaster. I. A Golgi analysis of wild-type structure. Cell Tissue Res. **258**, 441–475 (1989).
21. Song, S., Sjöström, P. J., Reigl, M., Nelson, S. & Chklovskii, D. B. Highly nonrandom features of synaptic connectivity in local cortical circuits. PLoS Biol. **3**, e68 (2005).
22. Varshney, L. R., Chen, B. L., Paniagua, E., Hall, D. H. & Chklovskii, D. B. Structural properties of the Caenorhabditis elegans neuronal network. PLOS Comput. Biol. **7**, e1001066 (2011).
23. Campos-Ortega, J. A. & Strausfeld, N. J. in Information Processing in the Visual Systems of Arthropods (ed. Wehner, R.) 31–36 (Springer Verlag, 1972).
24. Franceschini, N., Kirschfeld, K. & Minke, B. Fluorescence of photoreceptor cells observed in vivo. Science **213**, 1264–1267 (1981).
25. Douglass, J. K. & Strausfeld, N. J. Anatomical organization of retinotopic motion—sensitive pathways in the optic lobes of flies. Microsc. Res. Tech. **62**, 132–150 (2003).
26. Bausenwein, B., Dittrich, A. P. & Fischbach, K. F. The optic lobe of Drosophila melanogaster. II. Sorting of retinotopic pathways in the medulla. Cell Tissue Res. **267**, 17–28 (1992).
27. Bausenwein, B. & Fischbach, K. F. Activity labeling patterns in the medulla of Drosophila melanogaster caused by motion stimuli. Cell Tissue Res. **270**, 25–35 (1992).
28. Strausfeld, N. J. & Lee, J. K. Neuronal basis for parallel visual processing in the fly. Vis. Neurosci. **7**, 13–33 (1991).
29. Clark, D. A., Bursztyn, L., Horowitz, M. A., Schnitzer, M. J. & Clandinin, T. R. Defining the computational structure of the motion detector in Drosophila. Neuron **70**, 1165–1177 (2011).
30. Joesch, M., Schnell, B., Raghu, S. V., Reiff, D. F. & Borst, A. ON and OFF pathways in Drosophila motion vision. Nature **468**, 300–304 (2010).
31. Rister, J. et al. Dissection of the peripheral motion channel in the visual system of Drosophila melanogaster. Neuron **56**, 155–170 (2007).

32. Schnell, B., Raghu, S. V., Nern, A. & Borst, A. Columnar cells necessary for motion responses of wide-field visual interneurons in *Drosophila*. *J. Comp. Physiol. A* **198,** 389–395 (2012).
33. Tuthill, J. C., Nern, A., Rubin, G. M. & Reiser, M. B. Contributions of the 12 neuron classes in the fly lamina to motion vision. *Neuron* **79,** 128–140 (2013).
34. Douglass, J. K. & Strausfeld, N. J. Visual motion-detection circuits in flies: parallel direction-and non-direction-sensitive pathways between the medulla and lobula plate. *J. Neurosci.* **16,** 4551–4562 (1996).
35. Maisak, M. S. *et al.* A directional tuning map of *Drosophila* elementary motion detectors. *Nature* http://dx.doi.org/10.1038/nature12320 (this issue).
36. Srinivasan, M. & Dvorak, D. Spatial processing of visual information in the movement-detecting pathway of the fly. *J. Comp. Physiol. A* **140,** 1–23 (1980).
37. Haag, J. & Borst, A. Recurrent network interactions underlying flow-field selectivity of visual interneurons. *J. Neurosci.* **21,** 5685–5692 (2001).
38. Gouwens, N. W. & Wilson, R. I. Signal propagation in *Drosophila* central neurons. *J. Neurosci.* **29,** 6239–6249 (2009).
39. Ashmore, J. F. & Copenhagen, D. R. Different postsynaptic events in two types of retinal bipolar cell. *Nature* **288,** 84–86 (1980).
40. Mizunami, M. Synaptic rectification model equivalent to the correlation-type movement detector. *Biol. Cybern.* **64,** 1–6 (1990).
41. Oyster, C. W. & Barlow, H. B. Direction-selective units in rabbit retina: distribution of preferred directions. *Science* **155,** 841–842 (1967).
42. Kim, I.-J., Zhang, Y., Yamagata, M., Meister, M. & Sanes, J. R. Molecular identification of a retinal cell type that responds to upward motion. *Nature* **452,** 478–482 (2008).
43. Euler, T., Detwiler, P. B. & Denk, W. Directionally selective calcium signals in dendrites of starburst amacrine cells. *Nature* **418,** 845–852 (2002).
44. Briggman, K. L., Helmstaedter, M. & Denk, W. Wiring specificity in the direction-selectivity circuit of the retina. *Nature* **471,** 183–188 (2011).
45. Meinertzhagen, I. & Hanson, T. in *The Development of* Drosophila Melanogaster Vol. 2 (eds Bate, M. & Martinez Arias, A.) 1363–1491 (Cold Spring Harbor Laboratory Press, 1993).
46. Blondel, V. D., Guillaume, J. L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008,** P10008 (2008).
47. Eck, N. & Waltman, L. VOS: a new method for visualizing similarities between objects. *Adv. Data Anal.* 299–306 (2007).

METHODS

Tissue preparation, electron microscopy and imaging. The right part of the brain of a wild-type Oregon R female fly was serially sectioned into 40-nm slices. A total of 1,769 sections, traversing the medulla and downstream neuropils (Fig. 1c), were imaged at a magnification of ×5,000. This process is detailed in the Supplementary Methods.

Semi-automated reconstruction pipeline. To obtain a dense electron microscopy reconstruction of the reference column, we used a sequence of automated alignment and segmentation steps, followed by manual proofreading and reconstruction, which we described as the semi-automated reconstruction pipeline[19].

Image alignment. We first found a rough alignment of the full image stack, ignoring artefacts such as folds, tears and dirt occlusions, by using TrakEM2 rigid registration[48] to align image blocks consisting of 20 sections of 9 × 9 mosaics, and then aligning blocks by an automated search over images at the centre of each mosaic. This rough alignment served to determine which images overlapped, allowing more precise analysis of tissues with artefacts, and in particular, large folds. Pixels much darker than average were assumed to correspond to folds and were used to divide each image into two or more connected components, called patches. For each pair of overlapping patches, both within a section (along the boundary) and section-to-section, points of correspondence were found by correlation (~1 per 500,000 overlapping pixels). A least-squares fit of these points with regularization for scale and skew was then used to produce a global affine alignment of all patches. Examination of errors from this fit identified images for which the automatic division into patches was inaccurate, and these divisions were corrected manually. Once a satisfactory fit was obtained, each patch of each image was then slightly distorted to provide a best match to its neighbour(s) while still remaining close to the global affine. More details are available in the ref. 49.

Automatic image segmentation. In the next step, we partitioned the medulla region of interest within the aligned stack of greyscale images into subsets of pixels belonging to individual neurons. Given that the resolution of the transmission electron microscopy (TEM) data set is anisotropic, we developed a two-step process comprising 2D segmentation to identify cross-sections of neurons followed by linkage of these segments in 3D. No single algorithm was used on all data, because many different segmentation techniques were tried in parallel with proofreading efforts, and it was counter-productive to re-segment portions already corrected. A typical 2D segmentation step entailed creating boundary probability maps using morphological features[50–52] followed by boosted edge learning[53], mitochondria detection to reduce false boundaries[54], followed by watershed segmentation[55] and agglomerative clustering[56] using mean and median boundary values to create 2D segments. The 3D linkage step constructed a linkage graph of consecutive 2D segments in adjacent sections. Again, several methods were used, including simple metrics such as overlap, and machine-learning approaches that computed appropriate weights of features from previously proofread data. Further details of some of our automatic segmentation approaches can be found in previous publications[19,57]. Given that all segmentation algorithms make mistakes resulting from imaging artefacts and low z-resolution, and because manual correction of over-segmentations is easier than under-segmentations, we tuned our automatic algorithms to produce an over-segmented image volume. Furthermore, we preserved watershed regions called super-pixels to facilitate the manual correction of over-segmentation in the next step. We have released the latest (and we believe best) version of the segmentation code that we used (https://bitbucket.org/shivnaga/sstem), but caution the reader that even our best automatic segmentation required extensive manual proofreading, correction, and annotation (see below) to yield the results we report in this paper.

Proofreading and reconstruction. We next inspected the results of automatic segmentation, corrected remaining errors, and assigned synapses to the proofread cell arbors. Because this was time-consuming, we trained a group of professional editors, referred to as proofreaders, whose work was supervised by three experienced electron microscopists (Sh.T., Sa.T. and P.K.R.) (experts). Proofreaders and experts performed their tasks using a dedicated custom software tool, Raveler (D.J.O. *et al.*, manuscript in preparation). In total, these proofreading steps took ~12,940 person-hours (including 900 person-hours contributed by our experts). There were five key steps within the reconstruction procedure: (1) volume proofreading, (2) synapse annotation, (3) postsynaptic tracing, (4) anchor body refinement, and (5) selective sparse tracing, detailed in the Supplementary Methods.

Reliability of the wiring diagram. As introduced above, we assigned two proofreaders to each synapse, to increase the reliability of proofreading. Characterizing this, in 48.2% of cases both proofreaders were unable to identify a parent cell either because they could not trace the process confidently or because it left the medulla region of interest. In 44.0% of cases, both proofreaders traced the PSD to the same anchor body, and in 7.5% of cases one proofreader was unable to complete the tracing whereas the other traced the PSD to an anchor body (numbers extracted from Supplementary Table 1). However, only in very few cases

(0.23%) did the two proofreaders reach different anchor bodies. These numbers suggest that a large fraction of connections will be missed, following the two proofreader agreement process; however, all connections that are identified have a very high probability of being correct.

To assess our reconstruction quality further, we generated two connectomes from the dual proofreader results—an inclusive version that included connections found by either proofreader, and a consensus version in which connections were accepted only when both proofreaders agreed. Comparing these two connectomes was generally reassuring. Although the inclusive connectome has ~16% more connections, all the additional connections had only one synaptic contact. All connections with two or more synapses are present in both connectomes (Supplementary Table 1). We used the consensus connectome for all our analyses. However, the conclusions remain unchanged when using the inclusive connectome.

In general, the high rates of missed synaptic contacts in our proofreading was tolerable for our project because our intent was to study connections with several, parallel synaptic contacts. We could confirm that the connectome contains a large fraction of such strong connections by plotting the distribution of the number of contacts between connected pairs of cells.

We found a strongly heavy-tailed distribution of the total numbers of contacts for each connected pair both in the whole connectome (inset in Fig. 3b), and within the subset of sparsely-traced cell-types involved in motion detection (Supplementary Fig. 3). Given that the sizes of T-bars within the medulla are relatively uniform (A. McGregor *et al.*, unpublished observations) and the size of synaptic structures is thought to correlate with their physiological strength[58], we viewed the number of parallel synaptic contacts between two neurons as a proxy of synaptic weight. Furthermore, making an assumption that the probability of missing a synaptic contact during proofreading is uniform across all postsynaptic sites, we can estimate that the consensus connectome contains all connections with >5 synapses with a confidence level >95% (Supplementary Methods). Therefore, we believe our consensus connectome is both precise and comprehensive for strong connections.

Constructing a connectome module. In constructing a repeating module within the medulla connectome, all the members of each class of neurons within the reference column were identified, and the number of synapses from those neurons to all other neurons of the postsynaptic class was averaged over the presynaptic cells. For the ultraperiodic cell types, for example, Tm3, this could result in a fractional weight. Furthermore, because there are 1.5 Tm3 cells per column on average, we computed the synaptic weight by multiplying it by 1.5. These fractional weights provide the mean connection strength over different columns, because some columns have only a single Tm3, whereas others have two.

The directional summation that was used here was chosen because, in our reconstruction, we attempted to proofread each postsynaptic element to its associated neuron. However, we did not attempt to proofread every presynaptic site back to its parent neuron (as some such elements might derive from non-reference columns, which were not densely reconstructed).

This method was modified for the four connections in the motion detection circuit: L1 to Mi1, L1 to Tm3, Mi1 to T4, and Tm3 to T4. In these cases, the number of synaptic contacts from the densely reconstructed medulla connectome was replaced by the number of synaptic contacts identified during sparse tracing of these specific connections (Supplementary Table 3).

Computing the Mi1 and Tm3 receptive field components. We computed the Mi1 and Tm3 components of the receptive field for each T4, by multiplying the number of synaptic contacts from each L1 neuron to a single intermediate Mi1 or Tm3 neuron by the number of contacts from that intermediate cell to the T4, and then summing over all the Mi1 and Tm3 neurons that receive input from the same L1 (Fig. 4e). This multiplication is equivalent to counting the number of independent synaptic routes from each L1 to each T4, in which each route must use a different pair of the synaptic contacts between the L1 and the intermediate target cells, and the intermediate targets and the T4.

Monte Carlo error estimate. Our proofreading methodology results in very few false positive errors, but many false negative errors between neuron pairs (see earlier). Therefore, it is highly probable that the observed number of synaptic contacts (m) is a subset of a higher, true number of synaptic contacts between two neurons (n). Assuming that, for any connected pair of neurons each synaptic contact had an equal false negative probability, and using that probability in a binomial distribution, our goal was to estimate the posterior probability, $P(n|m)$ (that is, the probability of the true n given the observed m), for different values of n. To do this, because the previous distribution over n is unknown, we approximated this posterior probability by the likelihood, $P(m|n)$ (that is, probability of observing m given some value of the true n), for different values of n. We generated 1,000 different estimates of the true connection weight, for every connected neuron pair, by sampling from the set of possible n values with the computed

likelihood given each value of n. Neuron pairs that were disconnected because of high false negative error rates were also allowed to connect if their neurite profiles overlapped in 3D. The number of contacts was estimated using the same sampling method as before (but with m equal to zero). In this way, we generated 1,000 different connectivity matrices. The displacement was then computed for each matrix. The eigenvalues and eigenvectors of the covariance matrix over all displacement vectors were computed, and used to plot an ellipse along the eigenvector axes with 2σ confidence intervals (Fig. 5a). To extend this construction to the mean displacement over a group of neurons, 1,000 variations of the mean were computed, by first sampling from the displacements generated for each neuron, and then computing the mean for each set of samples. The ellipses were then computed, in the same way as before, over this set of 1,000 mean values (Fig. 5b, top).

Effect of limited reconstruction size. Because the spread of the dendritic arbors of the Tm3 cells connected to a given T4 is greater than that of the Mi1 cells, Tm3 cells are more likely to be partially reconstructed, and, hence, missing L1 inputs near the edges of our 19 column reconstruction. To provide a measure of the effect of this cut-off on each T4, we first visually inspected each Tm3 neuron, and classified it into one of four classes, depending on the percentage of the arbor that was missing from the 19 column reconstruction. Tm3 cells that were reconstructed fully received input from an average of six L1 neurons. In contrast, Tm3 cells in the fourth class, which had most of their arbor outside our region of interest, received input from an average of only two L1 neurons. By summing up the fraction of L1 inputs missing, weighted by the fraction of inputs provided by each Tm3 class to the given T4, we obtained an estimate of the weighted fraction of L1 inputs missing (through the Tm3 channel) for each T4 (the x axis in Fig. 5b, bottom). We also used this metric to justify the removal of T4 cells missing >15% of their weighted L1 inputs (through the Tm3 channel), in constructing the mean responses for T4 cells with the same output direction preferences (Fig. 5b, top).

Dendrite orientation. After proofreading, the T4 arborizations within M10 were skeletonized, starting from the centre of the thickest branch, adapting the method of ref. 59 with a modified weighting function. The axonal branch point was determined by Sh.T. so as to identify the dendritic part of the medulla arborization precisely. The local orientation around each dendritic node was computed (using three nodes in both directions around each node, and ignoring any nodes with fewer than three adjacent nodes in both directions). The dominant dendritic branch orientation, for each T4 (Fig. 6b), was computed by taking the fast fourier transform (MATLAB and Statistics Toolbox, version R2012b) of the distribution of orientations across nodes, and defined to be the phase of the fundamental mode component of the transformation. Assuming a normal distribution for the dominant orientation of the cells assigned to a layer (via their axon arbor depth),

we computed the probability of each untraced T4 lying within each cluster of cells, given its measured dominant orientation. The assignment of cell T4-4 to layer 1 was significant ($P < 0.05$), but the assignment of the other two cells was not. The colour map for the dendritic arbors (Fig. 6a and Supplementary Fig. 5) was chosen by centring colours on the average of the dendritic branch orientation for each cluster (Fig. 6b), and varying the colour continuously between clusters.

Data access. At publication the skeletons of all cells will be uploaded to http://neuromorpho.org. We will host a website (https://openwiki.janelia.org/wiki/display/flyem/Medulla+TEM+Reconstruction) that will allow the viewer to search the consensus connectome for connected neuron pairs. We will also provide the entire segmentation with synapse annotations, along with Raveler to open the data set, upon request. The requesting party will need to supply a hard drive.

48. Cardona, A. *et al.* TrakEM2 software for neural circuit reconstruction. *PLoS ONE* **7**, e38011 (2012).
49. Scheffer, L., Karsh, B. & Vitaladevuni, S. Automated alignment of imperfect EM images for neural reconstruction. Preprint at http://arXiv.org/abs/1304.6034 (2013).
50. Canny, J. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-8**, 679–698 (1986).
51. Martin, D. R., Fowlkes, C. C. & Malik, J. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Trans. Pattern Anal. Mach. Intell.* **26**, 530–549 (2004).
52. Soille, P. *Morphological Image Analysis: Principles and Applications* 2nd edn, 316 (Springer-Verlag New York, 2003).
53. Dollar, P., Tu, Z. & Belongie, S. Supervised learning of edges and object boundaries. *IEEE Comp. Soc. Conf. Comp. Vis. Pattern Rec.* **2**, 1964–1971 (2006).
54. Vitaladevuni, S., Mishchenko, Y., Genkin, A., Chklovskii, D. C. & Harris, K. M. Mitochondria detection in electron microscopy images. *Workshop Microscopic Image Anal. Appl. Biol.* http://www.miaab.org/miaab-2008-papers/21-miaab-2008-abstract-05.pdf (2008).
55. Vincent, L. & Soille, P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Trans. Pattern Anal. Mach. Intell.* **13**, 583–598 (1991).
56. Mohanta, P. P., Mukherjee, D. P. & Acton, S. T. Agglomerative clustering for image segmentation. *Int. Conf. Pattern Rec.* **1**, 664–667 (2002).
57. Vitaladevuni, S. N. & Basri, R. Co-clustering of image segments using convex optimization applied to EM neuronal reconstruction. *IEEE Comp. Soc. Conf. Comp. Vis. Patt. Rec.* http://dx.doi.org/10.1109/CVPR.2010.5539901 2203–2210 (2010).
58. Kasai, H., Matsuzaki, M., Noguchi, J., Yasumatsu, N. & Nakahara, H. Structure–stability–function relationships of dendritic spines. *Trends Neurosci.* **26**, 360–368 (2003).
59. Sato, M., Bitter, I., Bender, M. A., Kaufman, A. E. & Nakajima, M. TEASAR: Tree-structure extraction algorithm for accurate and robust skeletons. *Eighth Pac. Conf. Comp. Graphics Appl.* http://dx.doi.org/10.1109/PCCGA.2000.883951 281–287, 449 (2000).

LETTER

An observed correlation between plume activity and tidal stresses on Enceladus

M. M. Hedman[1], C. M. Gosmeyer[2], P. D. Nicholson[1], C. Sotin[3], R. H. Brown[4], R. N. Clark[5], K. H. Baines[3], B. J. Buratti[3] & M. R. Showalter[6]

Saturn's moon Enceladus emits a plume of water vapour and micrometre-sized ice particles from a series of warm fissures located near its south pole[1–10]. This geological activity could be powered or controlled by variations in the tidal stresses experienced by Enceladus as it moves around its slightly eccentric orbit. The specific mechanisms by which these varying stresses are converted into heat, however, are still being debated[11–16]. Furthermore, it has proved difficult to find a clear correlation between the predicted tidal forces and measured temporal variations in the plume's gas content[17–19] or the particle flux from individual sources[20,21]. Here we report that the plume's horizontally integrated brightness is several times greater when Enceladus is near the point in its eccentric orbit where it is furthest from Saturn (apocentre) than it is when near the point of closest approach to the planet (pericentre). More material therefore seems to be escaping from beneath Enceladus' surface at times when geophysical models predict its fissures should be under tension[12,15,16] and therefore may be wider open.

This analysis focuses on 252 images of the Enceladus plume at wavelengths of 0.88–1.56 µm obtained by the Cassini spacecraft's Visual and Infrared Mapping Spectrometer (VIMS)[22] between 2005 and 2012 (see Supplementary Information). Although these VIMS observations could not resolve individual jets and sources, they all had sufficient resolution and signal-to-noise ratio to detect the plume as a whole (see Fig. 1). The position of Enceladus along its orbit during each of these observations is given by the 'orbital phase', f: that is, the difference between the moon's orbital longitude and the longitude of its pericentre (also known as the moon's true anomaly). For the data considered here, f varies between $-40°$ and $+200°$. Hence the observations sample times when Enceladus was near the pericentre ($f \approx 0°$) and near the apocentre ($f \approx 180°$) of its eccentric orbit, and span a broad range of tidal stress states.

Measurements made at different orbital phases and different times can only be sensibly compared to one another if we also account for variations in the viewing geometry, especially the phase angle α (that is, the angle between the light rays incident on the plume and the scattered light rays that reach the camera). The micrometre-sized plume particles are most efficient at scattering light at large phase angles[6], so the plume will appear brighter when viewed at higher phase angles. Fortunately, the VIMS observations covered a range of phase angles when Enceladus was near the pericentre and the apocentre of its eccentric orbit, so we can control and compensate for these brightness variations due to changes in the viewing geometry. For example, the data shown in Fig. 1 compare measurements made at two different orbital phases for two different phase angles. At both phase angles the plume is brighter in the observation obtained when Enceladus was near the apocentre of its orbit, which strongly suggests that tidal forces do play an important role in controlling Enceladus' activity.

Owing to variations in the distance between Cassini and Enceladus, different observations sample the plume's brightness at different altitudes. Hence, in order to derive comparable quantitative estimates of the plume's brightness, we interpolate the brightness data from each observation to a common altitude. Fortunately, the plume's brightness decreases with altitude in a regular manner. Let us define the plume's 'equivalent width' (EW) at a given altitude z above Enceladus' south pole as the total integrated brightness in a horizontal slice through the plume (that is, a fixed z) after removing any background signal from Saturn's E ring (see Supplementary Information). For low-optical-depth systems like the plume, this quantity is insensitive to both the image resolution and the alignment of the fissures, which facilitates comparisons between different observations. Furthermore, as shown in Fig. 2, EW is a nearly linear function of the parameter Z ($= [z/(r_E + z)]^{1/2}$, where $r_E = 250$ km is the radius of Enceladus). This trend is not only a useful empirical fit to the data, it can also be physically justified on the basis of considerations of the velocity distribution of the plume particles[7]. For the observable parts of the plume, we can assume that the particle and gas density are so low that the particles follow purely ballistic trajectories and that Enceladus' gravity is by far the dominant force acting on the particles. In this situation, an individual particle launched from the surface at a velocity v_o that is less than Enceladus' escape speed, $v_{esc} = 240$ m s^{-1}, will reach an altitude z_{max} before it falls back to the surface. The velocity of the particle passes through zero when it reaches z_{max}, and energy conservation requires that $v_o = v_{esc}[z_{max}/(r_E + z_{max})]^{1/2}$. Hence a particle launched at speed v_o spends the most time near $Z = v_o/v_{esc}$ and makes the largest contribution to the plume's brightness at that location. Hence the steady decrease in the plume's brightness as a function of Z implies that fewer particles are launched at higher velocities, consistent with previous analyses[7].

Fitting the plume's EW profile in each image to a linear function of Z for altitudes between 50 and 450 km (that is, Z between 0.4 and 0.8), we may estimate EW_{85}, the plume's equivalent width at an altitude of 85 km (that is, $Z = 0.5$) for each observation. Figure 3 plots these estimates of the plume's brightness as a function of the observed phase angle. Note that at any given phase angle, observations taken when Enceladus was near its orbital apocentre are systematically brighter than those taken when Enceladus was near its orbital pericentre. (This trend persists even if we control for other geometric parameters like the sub-spacecraft latitude or longitude.) Furthermore, the data obtained when Enceladus was close to either its pericentre or its apocentre appear to follow a simple empirical phase function $P(\alpha)$ where the brightness is a power-law function of the scattering angle $\theta = 180° - \alpha$ with an index of around 2.5 (that is, $P(\alpha) \propto \theta^{-2.5}$, see Fig. 3). We may therefore define a 'corrected equivalent width', $CEW = EW_{85} \times [\theta/20°]^{+2.5}$. So long as the plume's phase function is approximately proportional to $\theta^{-2.5}$, then these corrected widths should be nearly independent of phase angle and

[1]Center for Radiophysics and Space Research, Cornell University, Ithaca, New York 14853, USA. [2]Department of Astronomy, Indiana University, Bloomington, Indiana 47405, USA. [3]Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, California 91109, USA. [4]Department of Planetary Sciences, Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA. [5]United States Geological Survey Mail Stop 964, Box 25046, Denver Federal Center, Denver, Colorado 80225, USA. [6]SETI Institute, 189 Bernardo Avenue Suite 100, Mountain View, California 94043, USA.
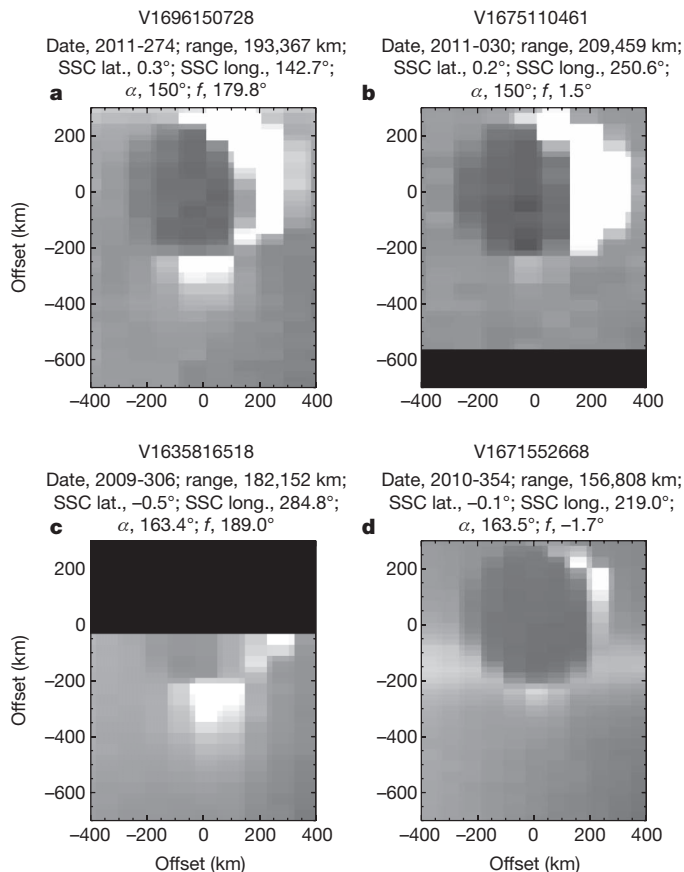
Figure 1 | Sample VIMS observations of Enceladus and its plume. All the images are made at wavelengths of 0.88–1.56 μm, have been rotated so that Enceladus' north pole points straight up, and have been projected onto a common spatial scale. In each panel, Enceladus appears as a dark disk silhouetted against the E-ring. The bright crescents at the upper right of these disks correspond to the illuminated part of the moon, and the plume appears as a diffuse bright streak below the moon's south pole (compare with Supplementary Fig. 2). Black regions correspond to regions not covered by the original observation. The observation name is given at the top of each panel, with the relevant observation date and geometric parameters (see Supplementary Information). Images **a** and **b** were obtained at a phase angle (α) of 150°, whereas images **c** and **d** were acquired at $\alpha \approx 163.5°$. Images taken at the same phase angle use the same stretch, but the images taken at a phase angle of 150° have a different stretch from the ones obtained at around 163.5° phase. Note that the plume is significantly brighter in **a** and **c** (where Enceladus was near its orbital apocentre), than it is in **b** and **d** (where the moon was near its orbital pericentre). Above each panel are given: date (year-day); range (distance between the spacecraft and Enceladus), sub-spacecraft latitude (SSC lat.); sub-spacecraft longitude (SSC long.); phase angle α; and orbital phase, f.

correspond to the plume brightnesses VIMS would have measured if it had always observed the plume at a phase angle of 160°.

Figure 4 shows a plot of the resulting corrected equivalent width estimates as a function of Enceladus' orbital phase. Note that the data from different phase angles follow the same trend, indicating that our approximate phase function is reasonably successful at correcting all these data. This plot confirms that the plume is indeed substantially brighter when Enceladus is at apocentre than when it is at pericentre. In fact, the plume's integrated brightness increases by more than a factor of three as Enceladus moves from pericentre to apocentre, and most of this change seems to occur between orbital phases of 90° and 180°.

Initial investigations of the longer-wavelength data obtained by VIMS reveal the same basic trends with orbital phase, albeit at lower signal-to-noise ratio. Thus far we have not detected any statistically significant variations in the shape of the plume's spectrum between observations taken when Enceladus is at pericentre versus apocentre,
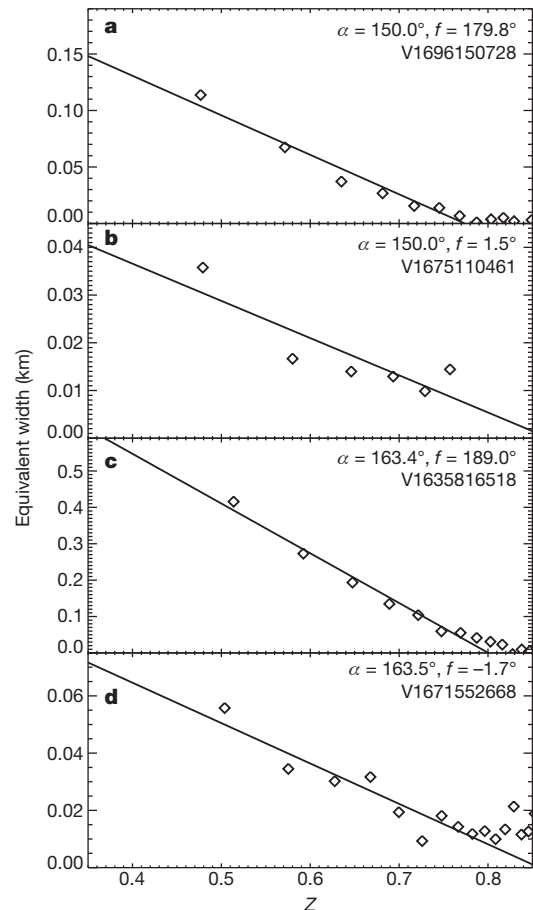
Figure 2 | Sample vertical profiles of the plume's brightness. a–d, The plume's equivalent width as a function of $Z = [z/(r_E + z)]^{1/2}$ for the four observations shown in Fig. 1a–d respectively. (As discussed in the text, Z can be regarded as a proxy for the particles' launch velocity.) In each panel, the diamonds show the measurements while the line shows a simple linear fit to the data between altitudes z of 50 and 450 km ($Z = 0.4$ and 0.8, respectively). This range was chosen because it excludes regions that are either too close to Enceladus (where the moon's limb may contaminate the signal) or too far from the moon (where the signal is weak). Note that in all four cases, this simple model provides a reasonable match to the trends in the data.

implying that the observable particle size distribution is not a strong function of orbital phase.

Similarly, the data taken between 2009 and 2012 all exhibit the same variations with orbital phase (see Fig. 4), indicating that the plume's activity level at a given orbital phase has not changed radically in the past few years. However, the 2005 observations yield brightness levels that are roughly 50% higher than comparable later observations. This may represent a decrease in the plume's average activity level between 2005 and 2009. However, even if this turns out to be the case, the 2005 data show the same trend of increasing brightness with increasing orbital phase as the later data. The variation in the plume's activity on orbital timescales therefore appears to be a persistent phenomenon.

These trends are also insensitive to altitude up to 300 km from the moon's surface (where the plume is clearly detectable). However, at higher altitudes, these trends might reverse owing to variations in the plume's scale height with orbital phase. In Fig. 2, the linear fits for the two observations made when Enceladus was close to its orbital apocentre intercept the x-axis when $Z \approx 0.8$. By contrast, the two observations made when Enceladus was near its orbital pericentre yield trends that go to zero when Z is at least 0.85. This distinction appears to be found consistently among the other observations: the weighted average of the x-intercepts for the profiles obtained when Enceladus was within $\pm 40°$ of pericentre occurs at $Z = 0.835 \pm 0.013$, whereas for the
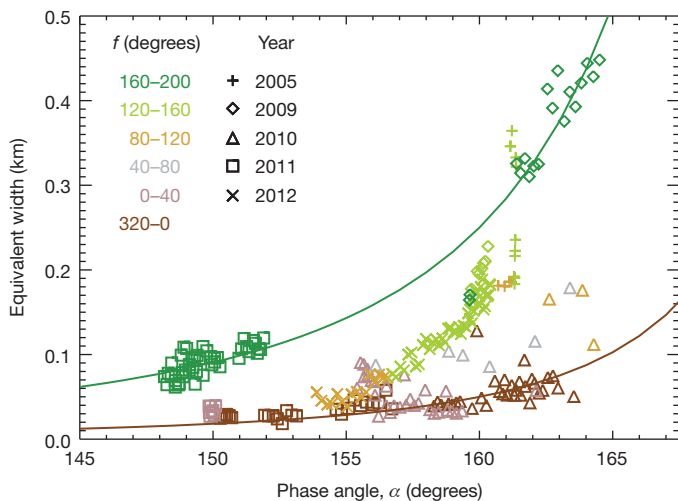
Figure 3 | Variations in the plume's brightness with phase angle. This plot shows the plume's equivalent width at 85 km altitude ($Z = 0.5$) and 0.88–1.56 μm as a function of phase angle α. Colours indicate observations made when Enceladus was at different orbital phases f, and symbols indicate when the observations were taken. The 1 s.d. statistical error bars on these data points are smaller than the symbol sizes (they range between 0.001 and 0.03, with most being between 0.005 and 0.01, see Supplementary Information). This plot shows that at all phase angles, the plume is consistently brighter when it is observed close to Enceladus' orbital apocentre. The two lines show fiducial phase functions that are proportional to $\theta^{-2.5}$, where θ is the scattering angle. The data obtained near Enceladus' apocentre follow this phase function fairly closely. For the pericentre data, the data do not match the predicted trend quite as well, but the above phase function is still an acceptable approximation to the true phase curve.

data obtained within $\pm20°$ of apocentre the average intercept occurs at $Z = 0.793 \pm 0.008$. This can be interpreted as a difference in the maximum launch velocity of the observed particles, with $v_{max} = 200 \pm 3$ m s^{-1} when Enceladus is near pericentre and $v_{max} = 190 \pm 2$ m s^{-1} when Enceladus is near apocentre. Hence the particles visible at 0.88–1.56 μm seem to be launched with a slightly larger maximum speed
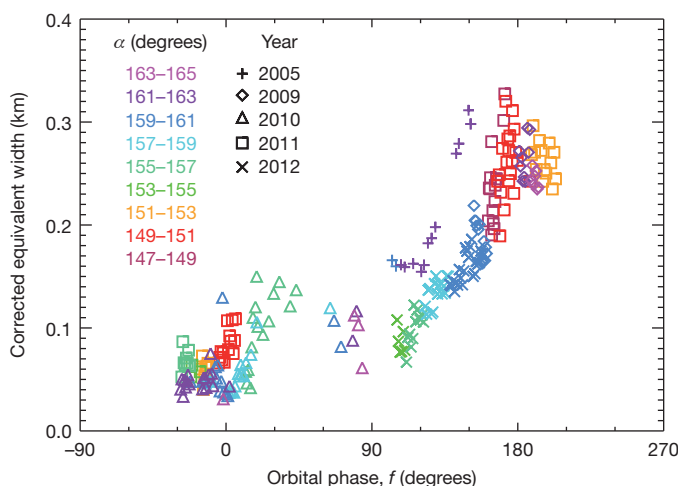
Figure 4 | Variations in the plume's corrected brightness with Enceladus' orbital position. This plot shows the plume's corrected equivalent width at 85 km altitude and 0.88–1.56 μm as a function of Enceladus' orbital phase f. Colours indicate observations made at different phase angles α, and symbols indicate when the observations were taken. The 1 s.d. statistical error bars on these data are smaller than the symbol sizes (they range between 0.001 and 0.03, with most being between 0.005 and 0.015, see Supplementary Information). Note that these data have now been corrected to remove the brightness variations due to varying phase angles by multiplying all the EW_{85} values by a factor of $(\theta/20°)^{2.5}$. With this correction applied, the data taken at different phase angles follow a common trend with orbital phase.

when Enceladus is near its orbital pericentre. This could reflect tidally induced changes in the crack geometry[23]. However, this trend is much more subtle than the variation in brightness, and so additional work will be needed before we can securely interpret this phenomenon.

A peak in plume activity when Enceladus is near its orbital apocentre is consistent with various geophysical calculations that suggest the normal stresses in Enceladus' south polar terrain will place the fissures under tension when Enceladus is near apocentre, and in compression when Enceladus is near pericentre[11,12,15,16,21]. Hence the data we report here provide strong evidence that tidal forces do play an important role in controlling Enceladus' plume activity, perhaps by changing the width of the conduits between the surface and various underground reservoirs.

1. Spencer, J. R. et al. in Saturn from Cassini-Huygens (eds Dougherty, M. K., Esposito, L. W. & Krimigis, S. M.) 683–724 (Springer, 2009).
2. Dougherty, M. K. et al. Identification of a dynamic atmosphere at Enceladus with the Cassini Magnetometer. Science 311, 1406–1409 (2006).
3. Hansen, C. J. et al. Enceladus' water vapor plume. Science 311, 1422–1425 (2006).
4. Waite, J. H. et al. Cassini Ion and Neutral Mass Spectrometer: Enceladus plume composition and structure. Science 311, 1419–1422 (2006).
5. Porco, C. C. et al. Cassini observes the active south pole of Enceladus. Science 311, 1393–1401 (2006).
6. Ingersoll, A. P. & Ewald, S. P. Total particulate mass in Enceladus plumes and mass of Saturn's E ring inferred from Cassini ISS images. Icarus 216, 492–506 (2011).
7. Hedman, M. M. et al. Spectral observations of the Enceladus plume with Cassini-VIMS. Astrophys. J. 693, 1749–1762 (2009).
8. Postberg, F. et al. The E-ring in the vicinity of Enceladus. Icarus 193, 438–454 (2008).
9. Postberg, F., Schmidt, J., Hillier, J., Kempf, S. & Srama, R. A salt-water reservoir as the source of a compositionally stratified plume on Enceladus. Nature 474, 620–622 (2011).
10. Spencer, J. R. et al. Cassini encounters Enceladus: background and the discovery of a south polar hot spot. Science 311, 1401–1405 (2006).
11. Nimmo, F., Spencer, J. R., Pappalardo, R. T. & Mullen, M. E. Shear heating as the origin of the plumes and heat flux on Enceladus. Nature 447, 289–291 (2007).
12. Hurford, T. A., Helfenstein, P., Hoppa, G. V., Greenberg, R. & Bills, B. G. Eruptions arising from tidally controlled periodic openings of rifts on Enceladus. Nature 447, 292–294 (2007).
13. Meyer, J. & Wisdom, J. Tidal heating in Enceladus. Icarus 188, 535–539 (2007).
14. Tobie, G., Cadek, O. & Sotin, C. Solid tidal friction above a liquid water reservoir as the origin of the south pole hotspot on Enceladus. Icarus 196, 642–652 (2008).
15. Smith-Konter, B. & Pappalardo, R. T. Tidally driven stress accumulation and shear failure of Enceladus's tiger stripes. Icarus 198, 435–451 (2008).
16. Hurford, T. A. et al. Geological implications of a physical libration on Enceladus. Icarus 203, 541–552 (2009).
17. Saur, J. et al. Evidence for temporal variability of Enceladus' gas jets: modeling of Cassini observations. Geophys. Res. Lett. 35, L20105 (2008).
18. Smith, H. T. et al. Enceladus plume variability and the neutral gas densities in Saturn's magnetosphere. J. Geophys. Res. 115, A10252 (2010).
19. Hansen, C. J. et al. The composition and structure of the Enceladus plume. Geophys. Res. Lett. 38, L11202 (2011).
20. Porco, C. et al. Jetting activity and thermal emission across the south polar terrain of Enceladus: observations and comparisons with shear-heating models. AGU Fall Meet. Abstr. P13F-02 (2011).
21. Hurford, T. A., Helfenstein, P. & Spitale, J. N. Tidal control of jet eruptions on Enceladus as observed by Cassini ISS between 2005 and 2007. Icarus 220, 896–903 (2012).
22. Brown, R. H. et al. The Cassini Visual and Infrared Mapping Spectrometer (VIMS) investigation. Space Sci. Rev. 115, 111–168 (2004).
23. Schmidt, J., Brilliantov, N., Spahn, F. & Kempf, S. Slow dust in Enceladus' plume from condensation and wall collisions in tiger stripe fractures. Nature 451, 685–688 (2008).

Author Contributions M.M.H., C.M.G. and P.D.N. performed the data analysis. R.N.C., along with P.D.N., C.S., R.H.B., K.H.B. and B.J.B., planned the observations, and provided the data from the VIMS instrument and the appropriate calibration routines. M.R.S. supported the photometric interpretation of these data. All authors discussed these results.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.M.H. (mmhedman@astro.cornell.edu).

LETTER

Squeezed light from a silicon micromechanical resonator

Amir H. Safavi-Naeini[1,2]*, Simon Gröblacher[1,2]*, Jeff T. Hill[1,2]*, Jasper Chan[1], Markus Aspelmeyer[3] & Oskar Painter[1,2,4]

Monitoring a mechanical object's motion, even with the gentle touch of light, fundamentally alters its dynamics. The experimental manifestation of this basic principle of quantum mechanics, its link to the quantum nature of light and the extension of quantum measurement to the macroscopic realm have all received extensive attention over the past half-century[1,2]. The use of squeezed light, with quantum fluctuations below that of the vacuum field, was proposed nearly three decades ago[3] as a means of reducing the optical read-out noise in precision force measurements. Conversely, it has also been proposed that a continuous measurement of a mirror's position with light may itself give rise to squeezed light[4,5]. Such squeezed-light generation has recently been demonstrated in a system of ultracold gas-phase atoms[6] whose centre-of-mass motion is analogous to the motion of a mirror. Here we describe the continuous position measurement of a solid-state, optomechanical system fabricated from a silicon microchip and comprising a micromechanical resonator coupled to a nanophotonic cavity. Laser light sent into the cavity is used to measure the fluctuations in the position of the mechanical resonator at a measurement rate comparable to its resonance frequency and greater than its thermal decoherence rate. Despite the mechanical resonator's highly excited thermal state (10^4 phonons), we observe, through homodyne detection, squeezing of the reflected light's fluctuation spectrum at a level 4.5 ± 0.2 per cent below that of vacuum noise over a bandwidth of a few megahertz around the mechanical resonance frequency of 28 megahertz. With further device improvements, on-chip squeezing at significant levels should be possible, making such integrated microscale devices well suited for precision metrology applications.

The generation of states of light with fluctuations below that of vacuum has been of great theoretical interest since the 1970s[3,7–9]. Early experimental work demonstrated squeezing of a few per cent below the vacuum noise level in a large variety of different nonlinear systems, such as neutral atoms in a cavity[10], optical fibres[11] and crystals with bulk optical nonlinearities[12,13]. Modern experiments demonstrate squeezing of almost 13 dB (ref. 14). Initial research was mainly pursued as a strategy to mitigate the effects of shot-noise, the manifestation of vacuum noise in the intensity detection of light, given the possibility of improved optical communication[7] and better sensitivity in gravitational-wave detectors[3,8]. In recent years, in addition to being used in gravitational-wave detectors[15], squeezed light has enhanced metrology in more applied settings[16,17].

The vacuum fluctuations arising from the quantum nature of light determine our ability to resolve mechanical motion optically, and set limits on the perturbation caused by the act of measurement[18]. A system well suited to studying quantum measurement experimentally is that of cavity optomechanics, in which an optical cavity's resonance frequency is designed to be sensitive to the position of a mechanical system. By monitoring the phase and intensity of the reflected light from such a cavity, a continuous measurement of mechanical displacement can be made. Systems operating on this simple principle have been realized in a variety of experimental settings, such as in large-scale laser gravitational-wave interferometers[19], microwave circuits with

electromechanical elements[20], solid-state mechanical elements[21–23] and ultracold gas-phase atoms[6,24] integrated with or comprising Fabry–Pérot cavities, and on-chip nanophotonic cavities sensitive to mechanical deformations[25,26].

The canonical cavity-optomechanical system consists of an optical cavity resonance that is dispersively coupled to the position of a mechanical resonance. The Hamiltonian describing the interaction between light and mechanics is $H_{int} = \hbar g_0 \hat{a}^\dagger \hat{a} \hat{x}/x_{zpf}$, where $\hat{x} = x_{zpf}(\hat{b}^\dagger + \hat{b})$ is the mechanical position, $x_{zpf} = \sqrt{\hbar/2m_{eff}\omega_m}$ is the zero-point fluctuation amplitude, ω_m is the mechanical resonance frequency, m_{eff} is the effective motional mass of the resonator, g_0 is the frequency shift of the optical resonance for a mechanical amplitude of x_{zpf}, $\hbar$ is Planck's constant divided by 2π, $\hat{a}$ and $\hat{a}^\dagger$ are respectively the annihilation and creation operators for optical excitations, and $\hat{b}$ and $\hat{b}^\dagger$ are the analogous operators for mechanical excitations. The optical cavity decay rate, κ, is the loss rate of photons from the cavity and the rate at which optical vacuum fluctuations are coupled into the optical resonance[27]. Similarly, the mechanical damping rate, γ_i, is the rate at which thermal bath fluctuations couple to the mechanical system. In all experimental realizations of solid-state optomechanics so far, including that presented here, the optomechanical coupling rate, g_0, has been much smaller than κ. As such, without a strong coherent drive, the interaction of the vacuum fluctuations with the mechanics is negligible.

Under the effect of a coherent laser drive, the cavity is populated with a mean intracavity photon number $\langle n_c \rangle$, and we consider the optical fluctuations about the classical steady state, $\hat{a} \rightarrow \sqrt{\langle n_c \rangle} + \hat{a}$. This modifies the optomechanical interaction, resulting in a linear coupling between the fluctuations of the intracavity optical field, $\hat{X}_0 = \hat{a} + \hat{a}^\dagger$, and the position fluctuations of the mechanical system, $\hat{x}$: $H_{int} = \hbar G \hat{X}_0 \hat{x}/x_{zpf}$. The parametric linear coupling occurs at an effective interaction rate of $G \equiv \sqrt{\langle n_c \rangle} g_0$, and the mechanical motion is coupled to the intracavity optical field at a rate of $\Gamma_{meas} \equiv 4G^2/\kappa$. Through this interaction, the intensity fluctuations of the vacuum field, $\hat{X}_{\theta=0}^{(in)}(t)$, entering the cavity impart a force on the mechanical system:

$$\hat{F}_{BA}(t) = \frac{\hbar\sqrt{\Gamma_{meas}}}{x_{zpf}} \hat{X}_{\theta=0}^{(in)}(t) \tag{1}$$

This radiation-pressure shot-noise (RPSN) force has previously been measured in an ultracold atomic gas[24] and, more recently, on a macroscopic silicon nitride nanomembrane[28]. The mechanical motion is in turn recorded in the phase of the light leaving the cavity:

$$\hat{X}_\theta^{(out)}(t) = -\hat{X}_\theta^{(in)}(t) - 2\frac{\sqrt{\Gamma_{meas}}}{x_{zpf}}\hat{x}(t)\sin(\theta) \tag{2}$$

Here $\hat{X}_\theta^{(j)} = \hat{a}_j e^{-i\theta} + \hat{a}_j^\dagger e^{i\theta}$, $\hat{a}_{in}$ and $\hat{a}_{out}$ are respectively the operators of the input and reflected optical fields from the cavity, and θ is the quadrature angle with $\theta = 0$ and $\theta = \pi/2$ corresponding respectively to the intensity and phase quadratures. In such a measurement, the optical cavity has the role of the position detector, measuring the observable $\hat{x}$ at a rate Γ_{meas}, and the RPSN imposes a measurement back-action force on the mechanical system[2]. In addition to this back-action noise, thermal

fluctuations from the bath also drive the mechanical motion, with their magnitudes becoming comparable as Γ_{meas} approaches the thermalization rate, $\Gamma_{thermal}(\omega) \equiv \gamma_i \bar{n}(\omega)$, where $\bar{n}(\omega)$ is the thermal bath occupancy.

Formally, the output noise power spectral density (PSD) of the homodyne detector photocurrent, I, normalized to the shot-noise level is found by taking the Fourier transform of the autocorrelation of equation (2):

$$\bar{S}_{II}^{out}(\omega) = 1 + \frac{4\Gamma_{meas}}{x_{zpf}^2}\left[\bar{S}_{xx}\sin^2(\theta) + \frac{\hbar}{2}\text{Re}\{\chi_m\}\sin(2\theta)\right] \quad (3)$$

where $\bar{S}_{xx}(\omega)$ is the noise PSD of the mechanical position fluctuations of the resonator and $\chi_m(\omega) = (m_{eff}(\omega_m^2 - \omega^2 - i\gamma_i\omega_m))^{-1}$ is the mechanical susceptibility characterizing the response of the mechanical system to an applied force. The PSD $\bar{S}_{xx}(\omega)$ contains noise stemming from coupling to the thermal bath, quantum back-action noise from the light field and any other technical laser noise driving the mechanics (Supplementary Information). The three terms in $\bar{S}_{II}^{out}(\omega)$ in equation (3), from left to right, are due to shot-noise, mechanical position fluctuations and the cross-correlation between the back-action noise force and mechanical position fluctuations. Only the third term can have a negative noise PSD and give rise to squeezing.

The primary hurdle to observing such squeezing, as in many quantum measurements, is strongly coupling to a preferred detection channel while simultaneously minimizing unwanted environmental perturbations. Most relevant to the work presented here are frequencies detuned from resonance, $\omega_m > |\delta \equiv \omega_m - \omega| \gg \gamma_i$, for which the approximate output noise PSD is

$$\bar{S}_{II}^{out}(\omega > 0) \approx 1 + \frac{2\Gamma_{meas}}{\delta}\left[\frac{\omega_m}{\delta}\frac{\bar{n}(\omega)}{Q_m}(1 - \cos(2\theta)) + \sin(2\theta)\right]$$

where we have assumed that the mechanical position fluctuations are predominantly due to thermal bath coupling (at high optical power, back-action heating and laser noise also contribute to the mode occupancy). In this case, fluctuations from the thermal bath limit appreciable squeezing of the optical probe field to a regime in which $\bar{n}(\omega)/Q_m \lesssim 1$. This requirement is equivalent to having $Q_m\omega_m \gtrsim k_B T_b/\hbar$, where k_B is Boltzmann's constant and T_b is the bath temperature. Appreciable squeezing also requires a detuning $|\delta| \gtrsim (\bar{n}/Q_m)\omega_m$ and a corresponding measurement rate larger than this detuning. Therefore, in the presence of thermal bath noise, squeezing over a significant spectral bandwidth requires not only a large cooperativity ($C \equiv \Gamma_{meas}/\gamma_i$) between the optical and mechanical components, as realized in recent cooling experiments[20,26,29], but the more stringent requirement that the effective measurement back-action force be comparable to all forces acting on the mechanical resonator, including the elastic restoring force of the mechanical structure.

To meet the requirements of strong measurement and efficient detection, we designed a 'zipper-style' optomechanical cavity[25] with a novel integrated waveguide coupler fabricated from the 220-nm-thick silicon device layer of a silicon-on-insulator microchip (Fig. 1). The in-plane differential motion of the two beams at a fundamental frequency of $\omega_m/2\pi = 28$ MHz strongly modulates the co-localized fundamental optical resonance of the cavity with a theoretical vacuum coupling rate of $g_0/2\pi = 1$ MHz. As shown in Fig. 1b, we use a silicon waveguide with a high-reflectivity photonic crystal end-mirror to excite and collect light efficiently from the optical cavity. Light from the silicon waveguide is coupled to a single-mode optical fibre using an optical fibre taper and a combination of adiabatic mode coupling and transformation.

The experimental set-up used to characterize the zipper cavity system and measure the optomechanical squeezing of light is shown in Fig. 2a. The silicon sample is placed in a continuous-flow ^{4}He cryostat with a cold-finger temperature of 10 K. A narrowband laser beam is used to probe the optomechanical system and measure the mechanical

motion of the zipper cavity. A wavelength scan of the reflected signal from the cavity is plotted in Fig. 2b, showing an optical resonance with a linewidth $\kappa/2\pi = 3.42$ GHz at a wavelength of $\lambda_c = 1,540$ nm. Inefficiencies in the collection and detection of light result in additional uncorrelated shot-noise in the measured signal and can reduce the squeezing to undetectable levels. For the device studied here, the cavity coupling efficiency, corresponding to the percentage of photons sent into the cavity which are reflected, is determined to be $\eta_\kappa = 0.54$. The fibre-to-chip coupling efficiency is measured at $\eta_{CP} = 0.90$. A homodyne detection scheme allows for high-efficiency detection of arbitrary quadratures of the optical signal field. Characterization and optimization of the efficiency of the entire optical signal path and homodyne detection system yielded an overall set-up efficiency of $\eta_{set-up} = 0.48$, corresponding to a total signal detection efficiency of $\eta_{tot} = \eta_{set-up}\eta_\kappa = 0.26$.

Figure 2c shows the noise spectrum of the thermal motion of the mechanical resonator obtained by setting the laser frequency near the cavity resonance and tuning the relative local-oscillator phase of the homodyne detector, θ_{lock}, to measure the quadrature of the reflected signal in which mechanical motion is imprinted. The mechanical spectrum shows the in-plane differential mode of interest at $\omega_m/2\pi = 28$ MHz, as well as several other more weakly coupled mechanical resonances of the nanobeams and coupling waveguide. A high-resolution, narrowband spectrum of the in-plane differential mode is inset in Fig. 2c, and shows a linewidth of $\gamma_i/2\pi = 172$ Hz, corresponding to a mechanical Q-factor of $Q_m = 1.66 \times 10^5$. The vacuum coupling rate of the in-plane differential mode, measured from the detuning dependence of the optical spring shift and damping, is determined to be $g_0/2\pi = 750$ kHz, in good agreement with theory. From the calibration of the noise power under the Lorentzian distribution in Fig. 2c, the in-plane differential mode is found to thermalize at low optical probe power to a temperature of $T_b \approx 16$ K, corresponding to a thermal phonon occupancy of $\bar{n}(\omega_m) \approx 1.2 \times 10^4$. This yields a ratio
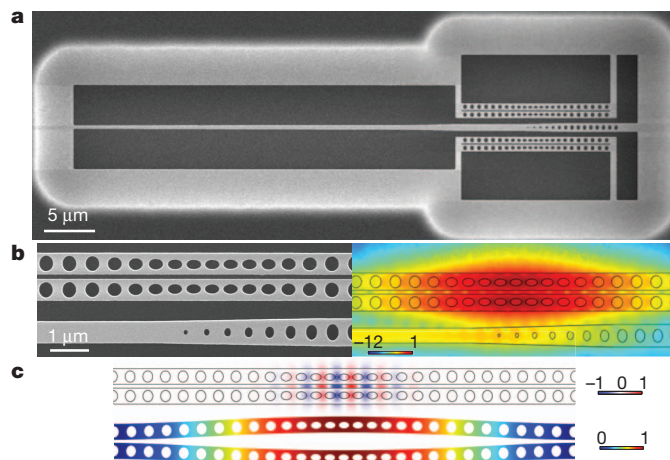

Figure 1 | Optomechanical device. a, Scanning electron microscope image of a waveguide-coupled zipper optomechanical cavity. The waveguide width is adiabatically tapered along its length and terminates with a photonic crystal mirror next to the cavity. The tapering of the waveguide allows for efficient input–output coupling and the photonic crystal termination makes the coupling to the cavity single sided. Two zipper cavities are coupled above and below the waveguide, each with a slightly different optical resonance frequency, allowing them to be separately addressed. **b**, Left: close-up of the coupling region between one of the cavities and the waveguide. Right: finite-element method (FEM) simulation of the cavity field leaking into the waveguide (log scale). Note that the field does not leak into the mirror region of the waveguide. **c**, Top: FEM simulation showing the in-plane electrical field of the fundamental optical cavity mode. Bottom: FEM simulation of the displacement of the fundamental in-plane differential mode of the structure with frequency $\omega_m/2\pi = 28$ MHz. The mechanical motion, modifying the gap between the beams, shifts the optical cavity frequency, leading to optomechanical coupling.
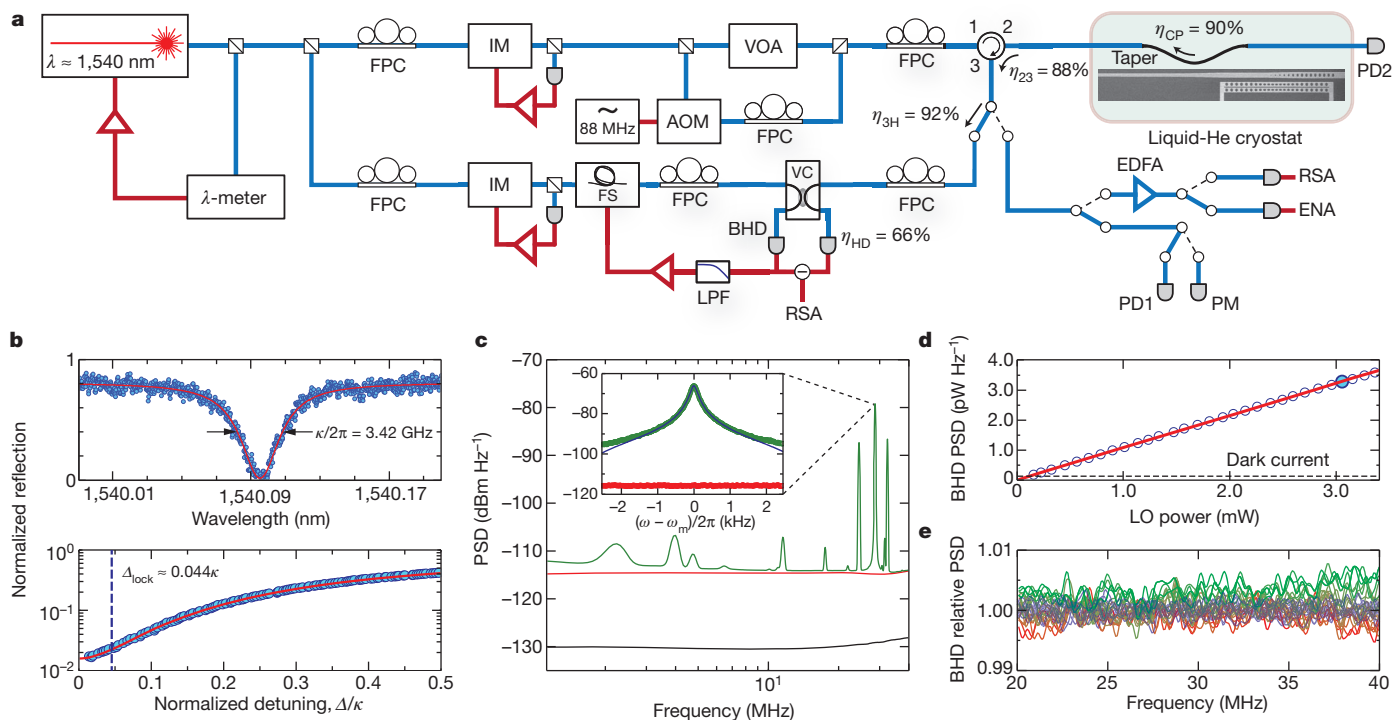
Figure 2 | Experimental set-up and device characterization. a, The optical signal is derived from an external-cavity diode laser and is sent into a tapered optical fibre inside a liquid-helium cryostat where the silicon sample is cooled to $T \approx 16\,\mathrm{K}$. The fibre taper is used to couple light evanescently into the silicon optomechanical device. The optical reflection from the device is collected by the same fibre taper and sent to a balanced homodyne detector (BHD) for characterization. For further details of the experimental set-up, see Methods Summary. The efficiencies of the circulator, switch and BHD are respectively denoted η_{23}, η_{3H} and η_{HD}. AOM, acousto-optic modulator; EDFA, erbium-doped fibre amplifier; ENA, electronic network analyser; FPC, fibre polarization controller; FS, fibre stretcher; IM, intensity modulator; λ-meter, wavemeter; LPF, low-pass filter; PD, photodetector; PM, power meter; RSA, real-time spectrum analyser; VC, variable coupler; VOA, variable optical attenuator. **b,** Top: reflected signal from the optical cavity at low optical power ($\langle n_c \rangle \approx 10$; linewidth, $\kappa/2\pi = 3.42\,\mathrm{GHz}$). Bottom: high-power ($\langle n_c \rangle \approx 790$)

reflected signal, showing the cavity-laser detuning (dashed line) locked to during squeezing measurements. **c,** Homodyne noise PSD of the reflected signal showing the transduced thermal Brownian motion of the zipper cavity at $T_b = 16\,\mathrm{K}$ (green curve; $\langle n_c \rangle = 80$). The red curve is the shot-noise level and the black curve is the detector's dark noise (in the absence of light input). Inset, close-up of the fundamental in-plane differential mechanical mode of the zipper cavity (fitted by blue curve; linewidth, $\gamma_i/2\pi = 172\,\mathrm{Hz}$). **d,** Mean value of the PSD of the BHD as a function of the local-oscillator (LO) power (signal blocked). The filled data point indicates the local-oscillator power used in the squeezing measurements. The red and dashed black curves correspond to a linear fit to the data and the level of the detector dark current, respectively. **e,** Noise PSD as a function of θ_{lock} (ranging from 0 (green) to π (red)) with the signal detuned far off resonance at $\Delta/\kappa \approx 30$, referenced to the noise level with the signal blocked (blue).

$\bar{n}(\omega_m)/Q_m \approx 0.072$, which is well within the regime where squeezing is possible.

To study accurately the noise properties of the reflected optical signal from the cavity, we make a series of measurements to characterize our laser and detection set-up. Figure 2d shows the measured noise PSD of the balanced homodyne detector for $\omega \approx \omega_m$ as a function of local-oscillator power, indicating a linear dependence on power and negligible ($\ll 0.1\%$) added noise above shot-noise. In the measured squeezing data discussed below, a local-oscillator power of 3 mW is used. Calibration of the laser intensity and frequency noise over the frequency range of interest ($\omega/2\pi = 1$–$40\,\mathrm{MHz}$) is described in Supplementary Information. The laser intensity noise is measured to be dominated by shot-noise over this frequency range, and the laser frequency noise is measured to be roughly flat in frequency at a level of $S_{\omega\omega} \approx 5 \times 10^3 \,\mathrm{rad^2\,Hz}$, which contributes insignificantly to the detected noise floor. Figure 2e shows the measured homodyne detector noise level normalized to vacuum noise for reflected laser light far-detuned from the optical cavity resonance. The small ($\sim \pm 0.15\%$) deviation in the measured noise level bounds the systematic uncertainty in the detector gain versus quadrature bias point as well as the contribution to the measured noise from optical elements other than the silicon cavity.

Measurements of the noise in the reflected optical signal from the cavity as a function of quadrature angle, frequency and signal power are shown in Figs 3 and 4. These measurements are performed for laser light on resonance with the optical cavity and for input signal powers

varying from 252 nW to 3.99 μW, with the maximum signal power corresponding to an average intracavity photon number of $\langle n_c \rangle = 3{,}153$. The laser is set at the appropriate cavity detuning for each signal power by scanning the wavelength across the cavity resonance while monitoring the reflection, and then stepping the laser frequency towards the cavity from the long-wavelength side until the reflection matches the level that corresponds to a detuning of $\Delta_{lock}/\kappa = 0.044 \pm 0.006$. This produces a shift of $\phi = (0.73 \pm 0.03)\pi$ in the measured quadrature angle from the on-resonance condition of equations (1)–(3) (Methods Summary).

In Fig. 3, we plot the theoretically predicted and measured noise PSDs versus quadrature angle for a signal power corresponding to $\langle n_c \rangle = 790$ photons. Each quadrature spectrum is the average of 150 traces, and after every other spectrum the signal arm is blocked and the shot-noise PSD is measured. The shot-noise level, corresponding to optical vacuum on the signal arm, is used to normalize the spectra. At certain quadrature angles, and for frequencies a few megahertz around the mechanical resonance frequency, we find that the light reflected from the zipper cavity shows a noise PSD below that of vacuum. The density plot of the theoretically predicted noise PSD (Fig. 3a) shows the expected wideband squeezing due to the strong optomechanical coupling in these devices, as well as a change in the phase angle where squeezing is observed below and above the mechanical frequency. This change is due to the change in sign of the mechanical susceptibility and the corresponding change in phase of the mechanical response to
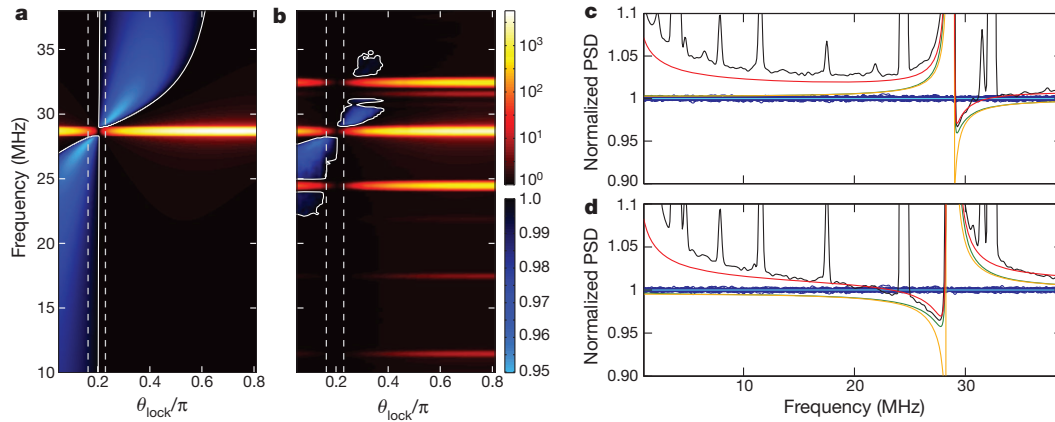
Figure 3 | Optomechanical squeezing of light. **a**, Theoretical model. Density plot of the predicted reflected-signal noise PSD, as measured on a balanced homodyne detector and normalized to shot-noise, for a simplified model of the optomechanical system (Supplementary Information). Areas with noise below shot-noise are shown in blue shades on a linear scale. Areas with noise above shot-noise are shown in orange shades on a log scale. The solid white line is a contour delineating noise above and below shot-noise. **b**, Experimental data. Density plot of the measured reflected-signal noise PSD for $\langle n_c \rangle = 790$, normalized to the measured shot-noise level. **c**, Slice of the measured density plot in **b** taken at $\theta_{lock}/\pi = 0.23$. **d**, Slice of the measured density plot in **b** taken at $\theta_{lock}/\pi = 0.16$. In **c** and **d**, the black curve corresponds to the measured data slice extracted from **b**. The dark blue traces are several measurements of the shot-noise level (average shown in light blue). Also shown is a model of the squeezing in the absence of thermal noise (orange curve), the same model with ohmic thermal noise of the mechanical mode included (green) and a full noise model including additional phenomenological noise sources (red curve). The vertical white dashed lines in **a** and **b** indicate the data slices shown in **c** and **d**.

RPSN. The measured noise PSD density plot (Fig. 3b) shows the presence of several other mechanical noise peaks and a reduced squeezing bandwidth, yet the overall phase- and frequency-dependent characteristics of the squeezing around the strongly coupled in-plane mechanical mode are clearly present. In particular, Fig. 3c and Fig. 3d show two slices of the noise PSD density plot in which the region of squeezing changes from being below the mechanical resonance frequency to being above it.

In Fig. 4a, we show the measured noise PSD as a function of quadrature angle for a frequency slice at $\omega/2\pi = 27.9$ MHz of the data shown in Fig. 3b. The measured squeezing and anti-squeezing are seen to be smaller and, respectively, larger than expected from a model of the optomechanical cavity without thermal noise. We also plot, in Fig. 4b, the maximum measured and modelled squeezing as functions of signal power. The simple theory predicts a squeezing level that monotonically increases with signal power, whereas the measured maximum squeezing saturates at a level of $4.5 \pm 0.2\%$ below the shot-noise at an intracavity power corresponding to $\langle n_c \rangle = 1,984$ photons. The error in the squeezing is dominated by the uncertainty in the linearity of the detector gain ($\pm 0.15\%$) and the variance of the measured shot-noise level ($\pm 0.1\%$).

To understand the processes that limit the bandwidth and magnitude of the measured squeezing, we plot in Fig. 4c the noise PSD for phase quadratures that maximize (left plot) and minimize (right plot) the transduction of the mechanical mode peak. Along with the measured data, we also plot the estimated noise due to phase noise of the signal laser, and that for a model of a single mechanical mode coupled to a thermal bath at $T_b = 16$ K. Low-frequency noise in the motion quadrature shows an ω^{-1} frequency dependence consistent with structural damping effects[30], but is much larger than that of the single-mode noise model. Noise in the quadrature that minimizes transduced motion is orders of magnitude larger than the noise predicted by the single-mode model and laser phase noise, and shows an $\omega^{-1/2}$ frequency dependence. The optical power dependence of the low-frequency noise in the motion quadrature indicates that optical absorption heats the structure to $T_b \approx 30$ K at the highest measured powers. As detailed in Supplementary Information, the red curve in each of the plots in Fig. 4 shows a full noise model incorporating structural damping noise from higher-frequency mechanical modes, optical absorption heating and a phenomenological $\omega^{-1/2}$ noise term.
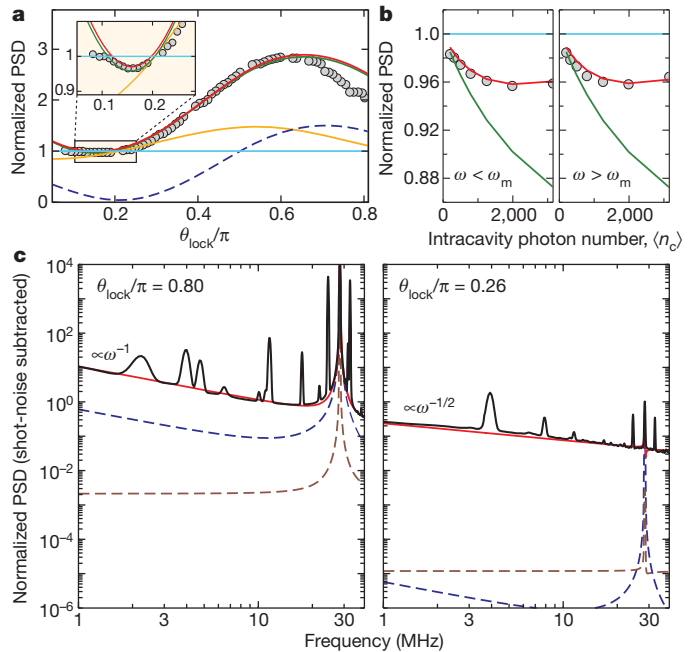

Figure 4 | Spectral and power dependence of noise. **a**, Measured balanced homodyne noise power of the reflected signal at $\omega/2\pi = 27.9$ MHz (filled circles) versus quadrature angle ($\Delta_{lock}/\kappa = 0.044$ and $\langle n_c \rangle = 790$). The green and red curves correspond to the single-mode and full noise models, respectively. The orange curve represents a model including the response of the mechanical mode in the absence of thermal noise, that is, when driven by RPSN only, and the dashed blue curve shows the thermal noise component. Inset, close-up of boxed region. **b**, Measured minimum noise PSD normalized to shot-noise (filled circles) versus $\langle n_c \rangle$. Left: maximum squeezing for $\omega < \omega_m$; right: maximum squeezing for $\omega > \omega_m$. Also shown are the single-mode noise model (green curve) and the full noise model (red curve). **c**, Balanced homodyne noise PSD of the reflected cavity signal for $\Delta_{lock}/\kappa = 0.052$ and $\langle n_c \rangle = 3,153$. Left: phase quadrature corresponding to maximum transduction of mechanical motion; right: phase quadrature corresponding to minimum transduction of mechanical motion. In each plot, the black curve is the measured data with the shot-noise level subtracted. Also shown are the modelled laser phase noise (dashed brown curve), the noise contribution from a single mechanical mode (dashed blue curve) and the full noise model (red curve).

These models indicate that the currently obtainable levels of squeezing are limited by the thermal noise of higher-order mechanical modes.

These measurements show that by reflecting light off a thin-film mechanical resonator undergoing large-amplitude thermal motion, light that is in certain respects 'quieter' than vacuum can be obtained. In contrast to previous work with ultracold gas-phase atoms[6], which used a narrowband atomic resonance and operated in short bursts owing to the atomic trap lifetime, the solid-state devices in this work allow for steady-state squeezing over almost 10 MHz of bandwidth and at optical frequencies completely tailorable through geometric design. The modest level of squeezing, limited by thermal noise and structural damping effects in the current devices, may also be substantially improved by increasing the mechanical Q-factor. Measurements of similar silicon devices with different surface treatments have yielded mechanical Q-factors as high as 7×10^5 (Supplementary Information), which, for the optical power levels used here, should enable on-chip squeezing 6 dB below shot-noise. Given the integrability of these microchip devices, optical extraction inefficiencies can be avoided by sending the squeezed light generated by one device directly into a second probe device. For example, such an on-chip squeezer and detector could form the basis of a quantum-enhanced micromechanical displacement and force sensor[17]. More generally, we expect these sorts of device to enable future experiments involving feedback and strong measurement of the dynamics of a mechanical system.

METHODS SUMMARY

Experimental set-up. A tunable external-cavity diode laser, actively locked to a wavemeter, is used to generate a strong local oscillator and the measurement signal. Fractions of the local oscillator and input signals are split off and detected, using intensity modulators to stabilize their power levels. Fibre polarization controllers adjust the polarization of the local oscillator and signal. A variable optical attenuator is used to set the signal power and an acousto-optic modulator is used to generate a tone for calibration (Supplementary Information). The reflected signal from the cavity is separated using a circulator, and is switched between one of three detection paths: one contains a power meter for power calibration, one contains a photodetector (PD1) for spectroscopy of the cavity and one contains an erbium-doped fibre amplifier for measurement of the mechanical spectrum on a real-time spectrum analyser or a network analyser. Squeezing of the reflected cavity signal is measured on a fourth path containing a variable coupler, where the signal is recombined with the local oscillator and detected on a balanced homodyne detector.

Homodyne phase angle. The relative phase between the local oscillator and the signal is determined from the low-pass-filtered component of the signal detected using a balanced homodyne detector, and set using a fibre stretcher. Locking the level of this signal determines the phase angle between the local oscillator and the reflected signal, which we call θ_{lock}. This angle differs from the phase θ between the light input to the cavity and the local oscillator, but is related to it through the phase response of the cavity: $\theta_{lock} = \theta - \phi$, where $\phi(\Delta) = \text{Arg}[1 - \kappa_e/(i\Delta + \kappa/2)]$ and κ_e is the extrinsic cavity coupling rate.

Cavity lock. A lock point slightly red detuned from resonance is chosen to avoid instabilities of the system resulting from thermo-optical effects. The laser is locked to this frequency using a wavemeter with a frequency resolution of $\pm 0.0015\kappa$. Drift of the optical cavity resonance over a single noise spectrum measurement (order of minutes) is found to be negligible. An estimate of the variance of Δ_{lock} is determined from the dependence of the transduction of the mechanical motion on the quadrature phase, which indicates that from one lock to another $\Delta_{lock}/\kappa = 0.044 \pm 0.006$.

1. Braginsky, V. & Khalili, F. *Quantum Measurements* (Cambridge Univ. Press, 1995).
2. Clerk, A. A., Devoret, M. H., Girvin, S. M., Marquardt, F. & Schoelkopf, R. J. Introduction to quantum noise, measurement, and amplification. *Rev. Mod. Phys.* **82**, 1155–1208 (2010).
3. Caves, C. M. Quantum-mechanical noise in an interferometer. *Phys. Rev. D* **23**, 1693–1708 (1981).
4. Fabre, C. *et al.* Quantum-noise reduction using a cavity with a movable mirror. *Phys. Rev. A* **49**, 1337–1343 (1994).
5. Mancini, S. & Tombesi, P. Quantum noise reduction by radiation pressure. *Phys. Rev. A* **49**, 4055–4065 (1994).
6. Brooks, D. W. C. *et al.* Non-classical light generated by quantum-noise-driven cavity optomechanics. *Nature* **488**, 476–480 (2012).
7. Yuen, H. P. Two-photon coherent states of the radiation field. *Phys. Rev. A* **13**, 2226–2243 (1976).
8. Hollenhorst, J. N. Quantum limits on resonant-mass gravitational-radiation detectors. *Phys. Rev. D* **19**, 1669–1679 (1979).
9. Walls, D. F. Squeezed states of light. *Nature* **306**, 141–146 (1983).
10. Slusher, R. E., Hollberg, L. W., Yurke, B., Mertz, J. C. & Valley, J. F. Observation of squeezed states generated by four-wave mixing in an optical cavity. *Phys. Rev. Lett.* **55**, 2409–2412 (1985).
11. Shelby, R. M., Levenson, M. D., Perlmutter, S. H., DeVoe, R. G. & Walls, D. F. Broadband parametric deamplification of quantum noise in an optical fiber. *Phys. Rev. Lett.* **57**, 691–694 (1986).
12. Wu, L.-A., Kimble, H. J., Hall, J. L. & Wu, H. Generation of squeezed states by parametric down conversion. *Phys. Rev. Lett.* **57**, 2520–2523 (1986).
13. Grangier, P., Slusher, R. E., Yurke, B. & LaPorta, A. Squeezed-light-enhanced polarization interferometer. *Phys. Rev. Lett.* **59**, 2153–2156 (1987).
14. Mehmet, M. *et al.* Squeezed light at 1550 nm with a quantum noise reduction of 12.3 dB. *Opt. Express* **19**, 25763–25772 (2011).
15. The LIGO Scientific Collaboration. A gravitational wave observatory operating beyond the quantum shot-noise limit. *Nature Phys.* **7**, 962–965 (2011).
16. Taylor, M. A. *et al.* Biological measurement beyond the quantum limit. *Nature Photon.* **7**, 229–233 (2013).
17. Hoff, U. B. *et al.* Quantum-enhanced micromechanical displacement sensitivity. *Opt. Lett.* **38**, 1413–1415 (2013).
18. Caves, C. M. Quantum-mechanical radiation-pressure fluctuations in an interferometer. *Phys. Rev. Lett.* **45**, 75–79 (1980).
19. Abbott, B. *et al.* Observation of a kilogram-scale oscillator near its quantum ground state. *N. J. Phys.* **11**, 073032 (2009).
20. Teufel, J. D. *et al.* Sideband cooling of micromechanical motion to the quantum ground state. *Nature* **475**, 359–363 (2011).
21. Gigan, S. *et al.* Self-cooling of a micromirror by radiation pressure. *Nature* **444**, 67–70 (2006).
22. Arcizet, O., Cohadon, P.-F., Briant, T., Pinard, M. & Heidmann, A. Radiation-pressure cooling and optomechanical instability of a micromirror. *Nature* **444**, 71–74 (2006).
23. Corbitt, T., Ottaway, D., Innerhofer, E., Pelc, J. & Mavalvala, N. Measurement of radiation-pressure-induced optomechanical dynamics in a suspended Fabry-Perot cavity. *Phys. Rev. A* **74**, 021802 (2006).
24. Murch, K. W., Moore, K. L., Gupta, S. & Stamper-Kurn, D. M. Observation of quantum-measurement backaction with an ultracold atomic gas. *Nature Phys.* **4**, 561–564 (2008).
25. Eichenfield, M., Camacho, R., Chan, J., Vahala, K. J. & Painter, O. A picogram- and nanometre-scale photonic-crystal optomechanical cavity. *Nature* **459**, 550–555 (2009).
26. Chan, J. *et al.* Laser cooling of a nanomechanical oscillator into its quantum ground state. *Nature* **478**, 89–92 (2011).
27. Gardiner, C. W. & Collett, M. J. Input and output in damped quantum systems: quantum stochastic differential equations and the master equation. *Phys. Rev. A* **31**, 3761–3774 (1985).
28. Purdy, T. P., Peterson, R. W. & Regal, C. A. Observation of radiation pressure shot noise on a macroscopic object. *Science* **339**, 801–804 (2013).
29. Verhagen, E., Deléglise, S., Weis, S., Schliesser, A. & Kippenberg, T. J. Quantum-coherent coupling of a mechanical oscillator to an optical cavity mode. *Nature* **482**, 63–67 (2012).
30. Liu, Y. T. & Thorne, K. S. Thermoelastic noise and homogeneous thermal noise in finite sized gravitational-wave test masses. *Phys. Rev. D* **62**, 122002 (2000).

Author Contributions A.H.S.-N., S.G. and M.A. designed the experiment. A.H.S.-N., S.G., J.C. and J.T.H. designed and fabricated the device, and performed the measurements. A.H.S.-N., S.G., J.T.H. and O.P. performed the analysis and modelling of the data. All authors were involved in writing and editing the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to O.P. (opainter@caltech.edu).

LETTER

Insolation–driven 100,000–year glacial cycles and hysteresis of ice–sheet volume

Ayako Abe-Ouchi[1,2,3], Fuyuki Saito[2], Kenji Kawamura[3,4], Maureen E. Raymo[5], Jun'ichi Okuno[1,2,3], Kunio Takahashi[2] & Heinz Blatter[1,6]

The growth and reduction of Northern Hemisphere ice sheets over the past million years is dominated by an approximately 100,000-year periodicity and a sawtooth pattern[1,2] (gradual growth and fast termination). Milankovitch theory proposes that summer insolation at high northern latitudes drives the glacial cycles[3], and statistical tests have demonstrated that the glacial cycles are indeed linked to eccentricity, obliquity and precession cycles[4,5]. Yet insolation alone cannot explain the strong 100,000-year cycle, suggesting that internal climatic feedbacks may also be at work[4–7]. Earlier conceptual models, for example, showed that glacial terminations are associated with the build-up of Northern Hemisphere 'excess ice'[5,8–10], but the physical mechanisms underpinning the 100,000-year cycle remain unclear. Here we show, using comprehensive climate and ice-sheet models, that insolation and internal feedbacks between the climate, the ice sheets and the lithosphere–asthenosphere system explain the 100,000-year periodicity. The responses of equilibrium states of ice sheets to summer insolation show hysteresis[11–13], with the shape and position of the hysteresis loop playing a key part in determining the periodicities of glacial cycles. The hysteresis loop of the North American ice sheet is such that after inception of the ice sheet, its mass balance remains mostly positive through several precession cycles, whose amplitudes decrease towards an eccentricity minimum. The larger the ice sheet grows and extends towards lower latitudes, the smaller is the insolation required to make the mass balance negative. Therefore, once a large ice sheet is established, a moderate increase in insolation is sufficient to trigger a negative mass balance, leading to an almost complete retreat of the ice sheet within several thousand years. This fast retreat is governed mainly by rapid ablation due to the lowered surface elevation resulting from delayed isostatic rebound[14–16], which is the lithosphere–asthenosphere response. Carbon dioxide is involved, but is not determinative, in the evolution of the 100,000-year glacial cycles.

Several internal feedback mechanisms have been suggested as crucial in 100-kyr glacial cycles, such as delayed bedrock rebound[14–16], the calving of ice-sheet margins[15], CO_2 variations[17,18], ocean feedback[16] and dust feedback[19,20]. The importance of these mechanisms needs to be investigated with physical models. Here we report numerical experiments with an ice-sheet model for the Northern Hemisphere, IcIES, in combination with the general circulation model (GCM) MIROC (Methods and Supplementary Fig. 1). Although it is not practical to run GCMs with fully coupled ice-sheet models on glacial–interglacial timescales[21], it is necessary to take into account the feedback from ice sheets on climate. In this study, a climate parameterization for the ice-sheet model is developed and calibrated using a suite of multi-snapshot atmospheric GCM experiments forced with different insolation values (for different eccentricities, obliquities and precessions), CO_2 concentrations and ice-sheet sizes, calculated in advance[22]. The ice-sheet model with the climate parameterization (IcIES–MIROC) can represent fast feedbacks, such as water vapour, cloud and sea-ice feedbacks,

and slow feedbacks, such as albedo/temperature/ice-sheet and lapse-rate/temperature/ice-sheet feedbacks[22]. We calculate the ice-sheet variation for the past 400 kyr forced by the insolation and atmospheric CO_2 content with improved dating[23] after running the simulation long enough to remove the dependence on the initial conditions (Figs 1a, b; Methods). After validating these results using palaeoclimate proxy data, we conducted sensitivity experiments to investigate the mechanism of ~100-kyr glacial cycles.

Our model realistically simulates the sawtooth characteristic of glacial cycles, the timing of the terminations and the amplitude of the Northern Hemisphere ice-volume variations (Fig. 1d) as well as their geographical patterns at the Last Glacial Maximum and the subsequent deglaciation (Supplementary Figs 2 and 3 and Supplementary Video 1). In the frequency domain, our model produces the largest spectral peak at a periodicity of ~100 kyr, as observed in the data (Fig. 1), even without the ocean feedback[16] or dust feedback[19]. In a series of model experiments, we investigated the roles of CO_2 (which also varies with a 100-kyr periodicity; Fig. 1b), various model parameters such as the time constant and the effective mantle density for isostatic rebound, and mass loss due to calving into proglacial lakes. The ~100-kyr periodicity, the sawtooth pattern and the timing of the terminations are reproduced with constant CO_2 levels[20,24] (for example 220 p.p.m.; Fig. 1e), and are robust for a range of model parameters (Supplementary Fig. 4).

By contrast, the spectral peak of ~100-kyr cycles is greatly reduced, and permanent large ice sheets remain, with the imposition of instantaneous isostatic rebound (Fig. 1f). This result supports the idea that the crucial mechanism for the ~100-kyr cycles is the delayed glacial isostatic rebound[14,15], which keeps the ice elevation low, and, therefore, the ice ablation high, while the ice sheet retreats. We note, however, that CO_2 variations can result in amplification of the full magnitude of ice-volume changes during the ~100-kyr cycles, but do not drive the cycles. Ice-sheet changes may induce variations in CO_2 through changing sea surface temperature, affecting the solubility of CO_2 (ref. 25), and through changing sea level, affecting the stratification of and CO_2 storage in the Southern Ocean[18]. During deglaciation, the melt water may affect ocean circulation, leading to an increase in atmospheric CO_2 (refs 23, 26, 27).

A striking feature of our results is that, in the experiments with constant CO_2 levels, the strong ~100-kyr cycle with a large amplitude appears only for the North American ice sheet within a particular range of CO_2 levels; the spectral peak of ~100-kyr cycle becomes small compared with those of ~41 and ~23-kyr cycles for CO_2 levels above 230 p.p.m. or below 190 p.p.m. (Fig. 1g). The Eurasian ice sheet responds only to insolation forcings at ~41-kyr and ~23-kyr periodicities, with small amplitudes in all cases (Fig. 1h). To investigate the mechanisms behind these observations, we conducted 200-kyr model experiments to obtain stable equilibria of both ice sheets for a range of prescribed climatic forcings, starting from either no ice or from large

ice sheets; we use summer temperature anomalies ranging from −5 to +3 K relative to the present day.

Because of strong albedo and topographic feedbacks, ice sheets are expected to have multiple stable equilibria[11–13]. We indeed observe two different equilibrium states for a range of climatic forcings, depending on the initial size of the ice sheets. Figure 2a shows maps of the equilibrium ice sheets and their corresponding surface mass balances, for various climate forcings, computed with large initial ice sheets. We also show the equilibrium volumes of the North American and Eurasian ice sheets versus the climate forcing (Fig. 2b), which both have hysteresis loops but with different shapes. For each ice sheet, the lower and upper branches in the ice-volume hysteresis loop (Fig. 2b, blue and red lines) correspond to equilibrium states resulting from small and large initial

states, respectively. The hysteresis branches define the ice-sheet states with neutral (equilibrium) mass balance for a given climatic forcing; the ice-sheet gains (or loses) mass if the climatic forcing falls below the lower branch (or rises above the upper branch). Crucially, the larger the ice sheet becomes, the smaller the forcing required for negative mass balance, as is reflected in the inclination of the upper branch. The positions and shapes of the hysteresis loops, and especially the inclinations of the upper branches, are quite different for the two ice sheets. The equilibrium states on the upper hysteresis branch of the North American ice sheet vary gradually over a wider range of forcings, from −2 to +2 K, relative to those of the Eurasian ice sheet, which range from −2 to −1 K (Fig. 2a, b).

To identify the physical mechanisms causing ~100-kyr cycles, we compare the results of equilibrium states with the simulated transient ice volume of the standard case with varying insolation and CO_2 forcings for the most recent glacial cycle (Fig. 1d); these data are plotted together in Fig. 2b. To enable the comparison, we converted insolation and CO_2 forcings to the summer temperature anomaly[22] (Methods). For the North American ice sheet, starting from the last interglacial forcing, 122 kyr before present (BP), with no ice, a rapid decrease in insolation well below the lower branch forces the mass balance to become positive and large, triggering the inception and growth of the ice sheet from the Canadian high Arctic, around latitude 70° N, to Labrador. Although the summer insolation maxima are large for the first two precessional cycles because of large eccentricity (104 and 84 kyr BP; Fig. 2b), the mass balance becomes negative only for a few thousand years because the upper hysteresis branch extends to high forcing values for small volumes.

As the ice sheet grows, the insolation forcing required for negative mass balance gradually becomes smaller. However, the reduction in eccentricity also makes the subsequent insolation maxima smaller, so the ice sheet continues to experience mostly a positive or near-neutral mass balance. By the fifth precession minimum (24 kyr BP) since the most recent interglacial period, near the eccentricity minimum, the volume of the North American ice sheet reaches nearly 90 m sea-level equivalent (that is, a volume equivalent to a change of 90 m in global sea level). At this stage, the southern margin of the large ice sheet is warm enough that a moderate climatic forcing can cause the ice sheet to retreat. With the subsequent increase in eccentricity, the summer insolation forcing in the next precessional cycle provides enough time and intensity for a rapid disintegration of the ice sheet (note the large excursion of insolation forcing above the upper hysteresis branch; Fig. 2b), which is why a large ice volume, called 'excess 100-kyr ice'[8], is observed before each glacial termination.

The upper branch in our hysteresis loop defines the threshold in ice volume and insolation at which the switch from a glacial state to a deglacial state occurs. Whereas the growth rate is governed by the gradual accumulation of snow, the retreat rate is governed by highly nonlinear processes such as the large ablation of ice that results from
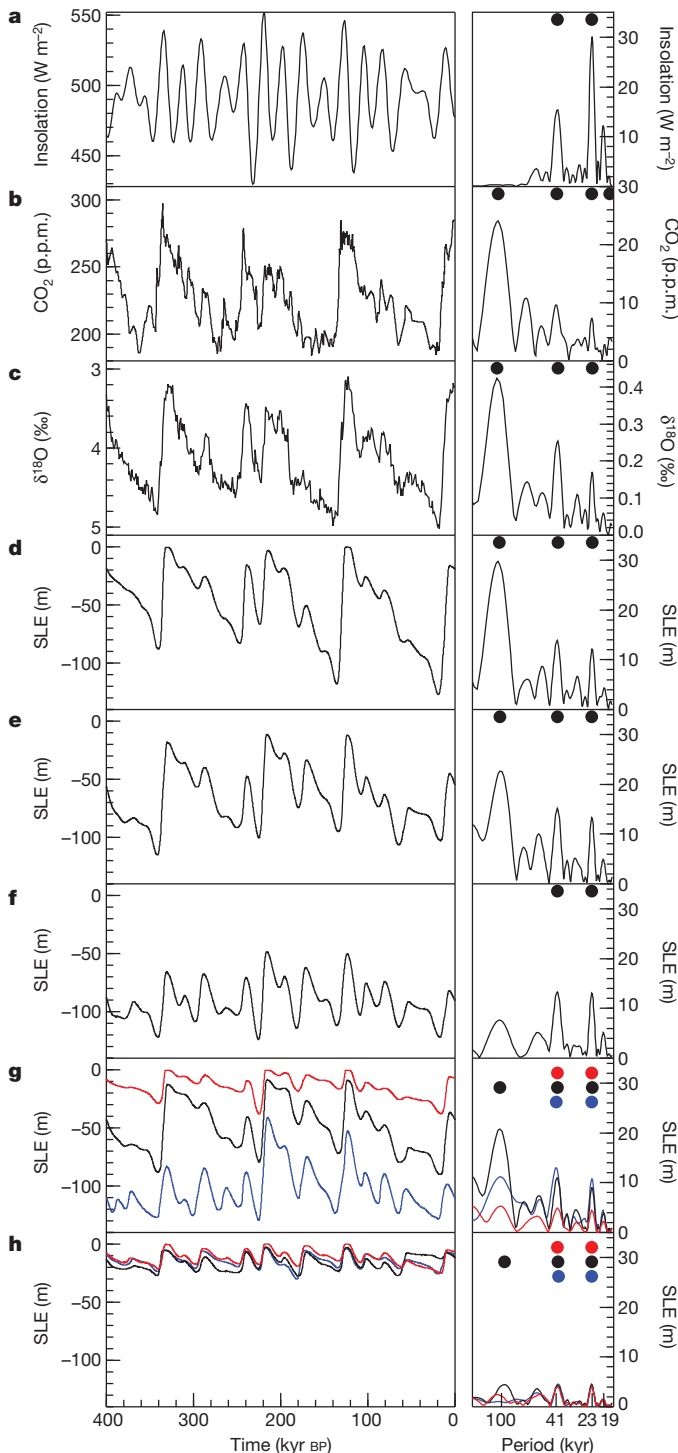
Figure 1 | Time series of forcing and responses of Northern Hemisphere ice sheets. Left, time series of the past 400 kyr; right, corresponding spectra. **a**, Mean extra-atmospheric insolation at latitude 65° N on 21 June of each year, which closely corresponds to the summer solstice. **b**, Atmospheric CO_2 from Vostok ice core on a revised timescale (ref. 23 and references therein). **c**, $\delta^{18}O$ from benthic foraminifera as a proxy for sea level and deep ocean temperature[30]. **d**, Modelled sea-level equivalent (SLE) of ice-volume changes relative to present with variations in atmospheric CO_2 content and insolation (standard case). **e**, Same as **d** but with a constant CO_2 concentration of 220 p.p.m. **f**, Same as **e** but with instant isostatic rebound. **g**, Same as **d** but with different constant CO_2 concentrations (blue, 160 p.p.m.; black, 220 p.p.m.; red, 260 p.p.m.) for the North American ice sheet. **h**, Same as **g** but for the Eurasian ice sheet. The spectra (right) show the amplitudes (calculated by the Multi-Taper Spectral Analysis Methods (MTM); using Analy Series; http://www.lsce.ipsl.fr/logiciels/index.php) in the corresponding frequencies of the time series (left). The coloured dots indicate peaks with more than 95% significance for the corresponding coloured curves.
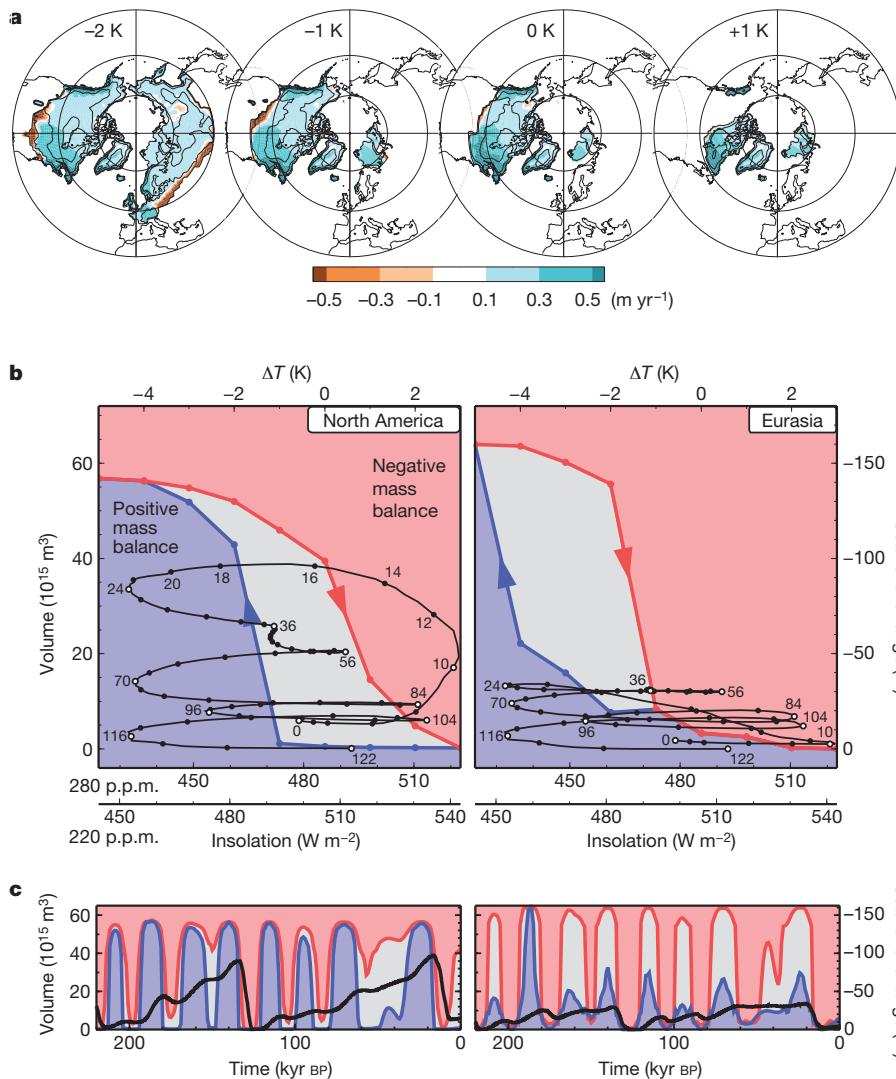
Figure 2 | Hysteresis of equilibrium states and transient evolution of the Northern Hemisphere ice sheets. a, Maps showing the equilibrium shapes and surface mass balances of ice sheets when the climatic anomalies relative to present conditions are respectively (left to right) −2, −1, 0 and 1 K (summer temperature) and when the model runs start from large initial ice sheets. Colours indicate the surface mass balance in metres per year. Note the large ablation areas and ablation rates (negative mass balance) that appear in the warm low latitudes. **b**, Modelled equilibrium and transient ice volumes as functions of the summer temperature anomaly for the North American (left) and Eurasian (right) ice sheets: red dots denote the large-volume equilibrium states if the model runs start from large initial ice sheets; blue dots show the small-volume equilibrium states for small initial ice sheets. The blue areas indicate a positive total mass balance of the ice sheet; red areas indicate a negative total mass balance. The black dots mark the evolution of the transient ice volume every 2 kyr for the last glacial cycle starting 122 kyr BP. The small numbers on the black trajectories show the corresponding time in kiloyears. The horizontal scales below the figures show the relation between the temperature anomaly (Methods) and the corresponding insolation at latitude 65° N on 21 June for two given constant atmospheric CO_2 concentrations (220 p.p.m. and 280 p.p.m.). **c**, Same as **b** but data shown as time series for the past two glacial cycles.

the low surface elevation due to the delayed isostatic response. Other processes may enhance the fast retreat, such as calving into proglacial lakes (Supplementary Fig. 2), increasing CO_2 concentrations, dust feedback[20], vegetation feedback[28] and basal sliding[29].

In contrast to the North American ice sheet, the dominant cycle of the volume of the Eurasian ice sheet has a period of ~40 kyr, and the volume never grows beyond 40 m sea-level equivalent. This pattern occurs for two reasons. First, the hysteresis branches of the ice volume are located within the lower half of the range of possible forcing variations (Fig. 2b). Thus, the ice sheet loses mass for a long time during insolation cycles. The difference in the positions of the hysteresis branches stems from the summers being generally warmer over Eurasia than over North America at high latitudes. Second, and more fundamentally, the upper hysteresis branch shows a step change, similar to that which occurs over Antarctica[13], whereby the volume decreases by 120 m sea-level equivalent for an increase in climatic forcing of only 1 K. Thus, regardless of the mean climatic state, the Eurasian ice sheet would not show ~100-kyr cycles because it cannot sustain intermediate ice volumes under the widely varying summer insolation forcing (equivalent to 6 K in a precessional cycle); the ice sheet can be only very large or very small (not shown). In summary, the shape and position of the hysteresis curve, different for each continent and for different constant CO_2 levels, are important in determining whether the dominant climatic cycle is ~100 kyr or ~40 kyr in period.

The ice sheets behave as a dynamical system: an ice sheet tends to approach a stable equilibrium state that also changes with time as the

climatic forcing changes. This behaviour is illustrated in Fig. 2c, which shows the time series of the volume evolution (black lines) and the attracting steady states (blue and red lines) corresponding to the hysteresis branches in Fig. 2b. Points where the stable equilibrium state lines cross correspond to changes in the sign of the mass balance and, thus, to changes between growing and shrinking ice sheets. The different timescales for growth (~10^4 yr) and decay (~10^3 yr) result in the decreases in volume evolution (Fig. 2c, black curves) to be much steeper than the increases. This asymmetry ultimately explains the characteristic sawtooth shape of the glacial cycles.

To understand the relative importance of the three astronomical parameters in generating the ~100-kyr cycles of the North American ice sheet, we conducted model experiments in which we kept fixed the eccentricity, obliquity or precession in turn, under a constant CO_2 concentration of 220 p.p.m. Results show that the ~100-kyr cycles persist for fixed obliquity, but not for fixed eccentricity or for fixed precession (Fig. 3 and Supplementary Fig. 6). These results demonstrate the essential role of precession and the eccentricity variation for the ~100-kyr cycle. Obliquity is not the driver of the ~100-kyr cycle, although it helps to amplify the ice-volume changes from glacial states to interglacial states. In summary, our model results suggest that the ~100-kyr cycle is essentially produced by the eccentricity modulation of precession amplitude through the changes in summer insolation[8], with the support of obliquity for glacial terminations, especially when eccentricity remains small after its minimum (for example at termination I 20–10 kyr BP and at termination IV 340–330 kyr BP).
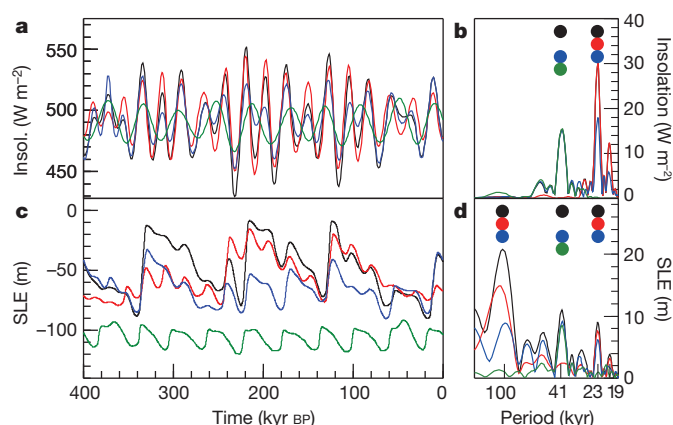
Figure 3 | Role of eccentricity, obliquity and precession in the 100-kyr cycle.
Time series of the model experiments with one of eccentricity, obliquity or precession fixed for a constant atmospheric CO_2 concentration of 220 p.p.m.
a, Insolation forcing (insolation at latitude $65°$ N on 21 June) with variations in eccentricity, obliquity and precession (black lines); with obliquity fixed at $23.5°$ (red lines); with eccentricity fixed at 0.02 (blue lines); and with perihelion passage fixed at the spring equinox and no precession (green lines).
b, Corresponding spectra of insolation change in **a** (as in Fig. 1a). **c**, Calculated ice-volume change, expressed as sea-level equivalent (colours same as in **a**).
d, Corresponding spectra of calculated ice-volume change in **c** (as in Fig. 1d).

A remarkable conclusion from our model results is therefore that the 100-kyr glacial cycle exists only because of the unique geographic and climatological setting of the North American ice sheet with respect to received insolation. Only for the North American ice sheet is the upper hysteresis branch moderately inclined; that is, there is a gradual change between large and small equilibrium ice-sheet volumes over a large range of insolation forcings. For this reason, as demonstrated in Fig. 2b, the amplitude modulation of summer insolation variation in the precessional cycle, due primarily to eccentricity, is able to generate the 100-kyr cycles with large amplitude, gradual growth and rapid terminations.

METHODS SUMMARY

A climate parameterization, taking into account the relevant climatic factors that control the ice-sheet evolution, is obtained from a suite of experiments using the MIROC GCM. On the basis of this climate parameterization, we drive the thermomechanically coupled shallow-ice-sheet model IcIES to study the impact of insolation and atmospheric CO_2 content on the change of Northern Hemisphere ice sheets. The experimental methods using the IcIES–MIROC model follow ref. 22 with a few modifications, as follows. The interaction between ice-sheet volume/area and the surface temperature is composed of the lapse-rate and albedo effects, and, as a novel term, the stationary-wave feedback, which is expressed as a spatial pattern of a temperature anomaly determined by GCM runs, weighted by a factor that depends on the ice-covered area over North America. The modifications in the ice sheet model concern the parameterization for basal sliding over the sediment and hard rock, which uses a realistic map of sediment thickness; the calving parameterization at the margin in terms of prescribed grounding-line flux; and the parameters in the isostatic rebound scheme, which are optimized by using a coupled ice-sheet/lithosphere/asthenosphere model.

Full Methods and any associated references are available in the online version of the paper.

Received 19 February; accepted 10 June 2013.

1. Clark, P. U. et al. The Last Glacial Maximum. Science 325, 710–714 (2009).
2. Hays, J. D., Imbrie, J. & Shackleton, N. J. Variations in Earth's orbit - pacemaker of ice ages. Science 194, 1121–1132 (1976).
3. Milankovitch, M. Kanon der Erdbestrahlung und seine Anwendung auf das Eiszeitproblem (R. Serbian Acad., 1941).
4. Lisiecki, L. E. Links between eccentricity forcing and the 100,000-year glacial cycle. Nature Geosci. 3, 349–352 (2010).
5. Huybers, P. Combined obliquity and precession pacing of late Pleistocene deglaciations. Nature 480, 229–232 (2011).
6. Saltzman, B., Hansen, A. R. & Maasch, K. A. The late Quaternary glaciations as the response of a three-component feedback system to Earth-orbital forcing. J. Atmos. Sci. 41, 3380–3389 (1984).
7. Tziperman, E., Raymo, M. E., Huybers, P. & Wunsch, C. Consequences of pacing the Pleistocene 100 kyr ice ages by nonlinear phase locking to Milankovitch forcing. Paleoceanography 21, PA4206 (2006).
8. Raymo, M. E. The timing of major climate terminations. Paleoceanography 12, 577–585 (1997).
9. Paillard, D. The timing of Pleistocene glaciations from a simple multiple-state climate model. Nature 391, 378–381 (1998).
10. Parrenin, F. & Paillard, D. Amplitude and phase of glacial cycles from a conceptual model. Earth Planet. Sci. Lett. 214, 243–250 (2003).
11. Abe-Ouchi, A. & Blatter, H. On the initiation of ice sheets. Ann. Glaciol. 18, 203–207 (1993).
12. Calov, R. & Ganopolski, A. Multistability and hysteresis in the climate-cryosphere system under orbital forcing. Geophys. Res. Lett. 32, L21717 (2005).
13. Pollard, D. & DeConto, R. M. Hysteresis in Cenozoic Antarctic ice-sheet variations. Global Planet. Change 45, 9–21 (2005).
14. Oerlemans, J. Model experiments on the 100,000-yr glacial cycle. Nature 287, 430–432 (1980).
15. Pollard, D. A simple ice-sheet model yields realistic 100 kyr glacial cycles. Nature 296, 334–338 (1982).
16. Deblonde, G., Peltier, W. R. & Hyde, W. T. Simulations of continental ice sheet growth over the last glacial-interglacial cycle: experiments with a one level seasonal energy balance model including seasonal ice albedo feedback. Palaeogeogr. Palaeoclimatol. Palaeoecol. 98, 37–55 (1992).
17. Shackleton, N. J. The 100,000-year ice-age cycle identified and found to lag temperature, carbon dioxide, and orbital eccentricity. Science 289, 1897–1902 (2000).
18. Paillard, D. & Parrenin, F. The Antarctic ice sheet and the triggering of deglaciations. Earth Planet. Sci. Lett. 227, 263–271 (2004).
19. Peltier, W. R. & Marshall, S. Coupled energy-balance ice-sheet model simulations of the glacial cycle: a possible connection between terminations and terrigenous dust. J. Geophys. Res. 100, 14269–14289 (1995).
20. Ganopolski, A. & Calov, R. The role of orbital forcing, carbon dioxide and regolith in 100 kyr glacial cycles. Clim. Past 7, 1415–1425 (2011).
21. Gregoire, L. J., Payne, A. J. & Valdes, P. J. Deglacial rapid sea level rises caused by ice-sheet saddle collapses. Nature 487, 219–222 (2012).
22. Abe-Ouchi, A., Segawa, T. & Saito, F. Climatic conditions for modelling the Northern Hemisphere ice sheets throughout the ice age cycle. Clim. Past 3, 423–438 (2007).
23. Kawamura, K. et al. Northern Hemisphere forcing of climatic cycles in Antarctica over the past 360,000 years. Nature 448, 912–916 (2007).
24. Berger, A., Loutre, M. F. & Gallee, H. Sensitivity of the LLN climate model to the astronomical and CO2 forcings over the last 200 ky. Clim. Dyn. 14, 615–629 (1998).
25. Kurahashi-Nakamura, T., Abe-Ouchi, A. & Yamanaka, Y. Effects of physical changes in the ocean on the atmospheric pCO2: glacial-interglacial cycles. Clim. Dyn. 35, 713–719 (2009).
26. Ganopolski, A. & Roche, D. M. On the nature of lead-lag relationships during glacial-interglacial climate transitions. Quat. Sci. Rev. 28, 3361–3378 (2009).
27. He, F. et al. Northern Hemisphere forcing of Southern Hemisphere climate during the last deglaciation. Nature 494, 81–85 (2013).
28. Claussen, M., Fohlmeister, J., Ganopolski, A. & Brovkin, V. Vegetation dynamics amplifies precessional forcing. Geophys. Res. Lett. 33, L09709 (2006).
29. Marshall, S. J. & Clark, P. U. Basal temperature evolution of North American ice sheets and implications for the 100-kyr cycle. Geophys. Res. Lett. 29, 2214 (2002).
30. Lisiecki, L. E. & Raymo, M. E. A Pliocene-Pleistocene stack of 57 globally distributed benthic δ18O records. Paleoceanography 20, PA1003 (2005).

Author Contributions A.A.-O. designed the research and experiments, and wrote the manuscript with F.S., K.K., M.E.R. and H.B. A.A.-O. and F.S. developed the numerical model, performed the experiments and analysed the results with K.T., K.K. and H.B. K.K. provided the ice-core data, and J.O. provided the Earth model for glacial isostatic rebound. All authors discussed the results and provided inputs on the manuscript.

METHODS

IcIES–MIROC model. The IcIES–MIROC model used in this study corresponds to the one described in ref. 22, with a few modifications explained below. The climate factors that control the ice-sheet changes, such as lapse rate and albedo feedback, are obtained from a suite of experiments using discrete GCM snapshots to obtain a climate parameterization[31]. On the basis of this climate parameterization, we drive the ice-sheet model to study the impact of orbital parameters and atmospheric CO_2 content on the change of Northern Hemisphere ice sheets.

Climate parameterization. To examine the response of climate to the orbital parameters, CO_2 and ice sheets, the atmospheric part of the atmosphere–ocean coupled GCM MIROC is used (K-1 developers, The University of Tokyo, 2004; http://ccsr.aori.u-tokyo.ac.jp/~hasumi/miroc_description.pdf). The model resolutions used in the present study are T106 (1° latitude, 1° longitude) and 20 vertical sigma levels with ~50-m thickness near the ice-sheet surface, and T42 with 11 levels, as in table 1 in ref. 22. The model includes dynamical and physical processes such as radiative transfer and high-resolution boundary-layer physics, which are necessary to resolve processes crucial for modelling the ice-sheet/climate system of the glacial cycles.

From the set of 18 sensitivity experiments with MIROC[22], including PMIP (Paleoclimate Modelling Intercomparison Project) experiments[32], the climatic effects of the changes in orbital parameters, atmospheric CO_2 content, lapse rate and surface albedo are separated and parameterized as follows:

$$T_s = T_{ref} + \Delta T_{sol} + \Delta T_{CO_2} + \Delta T_{ice} + \Delta T_{nonlinear}$$

Here T_s is the surface temperature and T_{ref} is a reference temperature based on the present-day climatology of the European Centre for Medium-Range Weather Forecasts (ECMWF)/ERA-40 meteorological re-analysis data (http://www.ecmwf.int/research/era/do/get/era-40). The terms ΔT_{CO_2} and ΔT_{sol} denote the changes in temperature according to changes in atmospheric CO_2 content and the change in temperature according to changes in insolation, respectively. The term $\Delta T_{nonlinear}$ is a residual term due to other feedback effects. The effects of the atmospheric response to changes in ice-sheet size, ΔT_{ice}, is decomposed into three terms

$$\Delta T_{ice} = \Delta T_{lapse} + \Delta T_{albedo} + \Delta T_{swf}$$

where the lapse-rate effect depends on the local surface elevation, the albedo effect depends on the ice-sheet size and the stationary-wave feedback of the atmosphere, ΔT_{swf}. The stationary-wave feedback is prescribed in the model runs by a temperature map, where the lapse-rate effect is subtracted from the difference between a model experiments with full ice-sheet topography and a model experiment with only flat ice, but both with ice albedo. It is expressed as a product of a spatial pattern of the temperature anomaly and a factor, r, that depends on the ice-covered area over North America:

$$r = \max\left[0, \min\left[1, \frac{A(t) - A_{12\,kyr}}{A_{max} - A_{12\,kyr}}\right]\right]$$

Here $A_{12\,kyr} = 8.146818 \times 10^{12}\ m^2$ and $A_{max} = 1.4 \times 10^{13}\ m^2$ are the assumed ice-covered area over North America 12 kyr BP and at the Last Glacial Maximum, respectively.

The parameterization of the effect of variable astronomical forcing and variable CO_2 follows that of ref. 22. For surface melt on the ice sheet, a positive-degree-day scheme following ref. 33 is applied. The astronomical forcing is based on ref. 34, and the CO2 forcing[35] is modified with revised dating[23].

IcIES ice-sheet model. The numerical ice-sheet model used in this study is the ice-sheet model for integrated Earth-system studies (IcIES), which is a thermomechanically coupled model in the shallow-ice approximation. The model is driven by surface boundary conditions such as the distributed temporal variations of climate in terms of surface mass balance and temperature, and by basal boundary conditions such as the bed topography, fixed geothermal heat flux and fixed sediment/hard-rock distribution. Sensitivity studies on model parameters and initial conditions

are shown in Supplementary Figs 2, 4 and 5. The model modifications compared with ref. 22 are as follows.

(1) Basal sliding. The parameterization for basal sliding over the sediment and hard-rock follows ref. 36. The grid points are categorized as either sediment type or hard-rock type. A sliding law of the form $u_b = CH|\nabla h|^n \nabla h$ is used, where u_b is the sliding velocity, H is the local ice thickness, ∇h is the surface inclination and C is the sliding coefficient. For the sediment-type grid points we use a linear sliding law with $C = 500\ yr^{-1}$ and $n = 0$, whereas for hard-rock-type grid points a nonlinear sliding law with $C = 10^5\ yr^{-1}$ and $n = 2$ is used. To prescribe the sediment area in the ice-sheet model, a global map of sediment thickness at a resolution of 1° by 1° provided by SEDMAP[37] is used. If the sediment thickness is more than 100 m, then sediment-type basal sliding is applied; otherwise, hard-rock-type sliding is applied.

(2) Calving. In addition to the passive calving at the margin of the land[22] (defined by a prescribed land mask), a parameterization of active calving[38] is implemented to represent a potential marine ice-sheet instability[39]. The calving flux (acting as grounding-line flux) at the margin is applied if a grid point satisfies the following three conditions: the bedrock elevation at the grid point is below sea level, corresponding to a marine ice-sheet situation; the surface mass balance at the grid point is negative, corresponding to an ablation area; and the grid faces the ocean, that is, at least one of the eight neighbouring grid points satisfies the floating condition. We apply a constant calving flux by replacing the surface ablation on this grid point by a fixed value ($-10\ m\ yr^{-1}$ in the standard run).

(3) Isostatic rebound. The dynamics of isostatic rebound is given by[22]

$$\frac{\partial b}{\partial t} = -\frac{1}{\tau}\left[b - b_0 + \frac{\rho_i}{\rho_{eff}}H\right]$$

where b, b_0, H, t and ρ_i are the transient bed elevation, the relaxed bed elevation without ice load, the ice thickness, time and the ice density, respectively. We introduce an effective mantle density of $\rho_{eff} = 4{,}500\ kg\ m^{-3}$ and a time constant of $\tau = 5{,}000\ yr$ in the present study (Supplementary Figs 2 and 4). The high effective density and the time constant are optimized using the viscoelastic Earth model of ref. 40 coupled to the IcIES ice-sheet model[41]. The high effective density compensates for the missing elastic forces in Earth's crust, which reduce the total isostatic motion[42].

31. Pollard, D. A retrospective look at coupled ice sheet-climate modeling. *Clim. Change* **100**, 173–194 (2010).
32. Braconnot, P. *et al.* Evaluation of climate models using palaeoclimatic data. *Nature Clim. Change* **2**, 417–424 (2012).
33. Reeh, N. Parameterization of melt rate and surface temperature on the Greenland ice sheet. *Polarforschung* **59**, 113–128 (1991).
34. Berger, A. L. Long-term variations of daily insolation and quaternary climatic changes. *J. Atmos. Sci.* **35**, 53–74 (1978).
35. Petit, J. R. *et al.* Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436 (1999).
36. Calov, R., Ganopolski, A., Petoukhov, V., Claussen, M. & Greve, R. Large-scale instabilities of the Laurentide ice sheet simulated in a fully coupled climate-system model. *Geophys. Res. Lett.* **29**, 2216 (2002).
37. Laske, G. & Masters, G. A global digital map of sediment thickness. *Eos Trans. AGU* **78**, F483 (1997).
38. Pollard, D. in *Milankovitch and Climate: Understanding the Response to Astronomical Forcing* Pt 2 (eds Berger, A., Imbrie, J., Hays, H., Kukla, G. & Saltzman, B.) 541–564 (Reidel, 1984).
39. Pollard, D. & DeConto, R. M. Modelling West Antarctic ice sheet growth and collapse through the past five million years. *Nature* **458**, 329–332 (2009).
40. Okuno, J. & Nakada, M. Effects of water load on geophysical signals due to glacial rebound and implications for mantle viscosity. *Earth Planets Space* **53**, 1121–1135 (2001).
41. Okada, Y. *Interaction between Northern Hemisphere Ice Sheet and Solid Earth Throughout the Ice Age Cycle* [in Japanese]. MSc thesis, Univ. Tokyo (2008).
42. Crucifix, M., Loutre, M. F., Lambeck, K. & Berger, A. Effect of isostatic rebound on modelled ice volume variations during the last 200 kyr. *Earth Planet. Sci. Lett.* **184**, 623–633 (2001).

LETTER

Nitrogen losses in anoxic marine sediments driven by *Thioploca*–anammox bacterial consortia

M. G. Prokopenko[1,2], M. B. Hirst[3], L. De Brabandere[4], D. J. P. Lawrence[5], W. M. Berelson[2], J. Granger[6], B. X. Chang[7], S. Dawson[3], E. J. Crane III[5], L. Chong[2], B. Thamdrup[4], A. Townsend-Small[8] & D. M. Sigman[7]

Ninety per cent of marine organic matter burial occurs in continental margin sediments, where a substantial fraction of organic carbon escapes oxidation and enters long-term geologic storage within sedimentary rocks. In such environments, microbial metabolism is limited by the diffusive supply of electron acceptors. One strategy to optimize energy yields in a resource-limited habitat is symbiotic metabolite exchange among microbial associations[1,2]. Thermodynamic and geochemical considerations indicate that microbial co-metabolisms are likely to play a critical part in sedimentary organic carbon cycling[3–5]. Yet only one association, between methanotrophic archaea and sulphate-reducing bacteria, has been demonstrated in marine sediments *in situ*[6,7], and little is known of the role of microbial symbiotic interactions in other sedimentary biogeochemical cycles[8]. Here we report *in situ* molecular and incubation-based evidence for a novel symbiotic consortium between two chemolithotrophic bacteria—anaerobic ammonium-oxidizing (anammox) bacteria and the nitrate-sequestering sulphur-oxidizing *Thioploca* species—in anoxic sediments of the Soledad basin at the Mexican Pacific margin. A mass balance of benthic solute fluxes and the corresponding nitrogen isotope composition of nitrate and ammonium fluxes indicate that anammox bacteria rely on *Thioploca* species for the supply of metabolic substrates and account for about 57 ± 21 per cent of the total benthic N_2 production. We show that *Thioploca*–anammox symbiosis intensifies benthic fixed nitrogen losses in anoxic sediments, bypassing diffusion-imposed limitations by efficiently coupling the carbon, nitrogen and sulphur cycles.

Interspecies metabolite exchange is an effective strategy for harvesting potential energy in resource-limited environments. In organic-rich marine sediments, which are characterized by a strong deficit of electron acceptors, symbiotic microbial consortia probably have a dominant role in organic carbon (C_{org}) cycling, given the capacity for such associations to maximize the energy yields of metabolic electron transfers[2–4]. Yet little is known of microbial co-metabolisms in the other biogeochemical cycles to which C_{org} transformations are linked[8]. The marine nitrogen (N) cycle is of particular interest in this respect because symbiotic microbial associations may affect the balance of N fluxes that, through a series of biogeochemical feedbacks, may regulate biological carbon cycling[9].

Nitrogen is supplied to the ocean through the fixation of atmospheric N_2 by diazotrophic bacteria. Fixed N is converted back to N_2 by 'denitrification' in the broad sense of the term: the enzymatic reduction of nitrate coupled to the oxidation of C_{org}, metals, methane, sulphide or ammonium (the last in a reaction known as anammox[10,11]) (Supplementary Information section 1.1). The balance between N_2 fixation and denitrification regulates the availability of N for marine photosynthesis. Symbioses between primary producers and diazotrophs support C_{org} production in severely N-limited regions of the ocean[12,13]. But no similar symbiotic associations have been demonstrated in denitrification, apart

from a tentative report of endobiotic denitrifying bacteria in allogromiid foraminifera[14].

In organic-rich suboxic/anoxic sediments, denitrification is expected to be limited by the diffusive flux of nitrate from the overlying water[15]. However, in sediments of low-oxygen (O_2) regions within the eastern Pacific (Fig. 1a), substantial N_2 production occurs at depths below the diffusion depth of nitrate[16–18]. From geochemical and isotopic observations, a symbiotic partnership was proposed between the large motile bacteria *Thioploca* and anammox bacteria, with the net effect of enhancing the conversion of fixed N to N_2 within anoxic sediments[16].

Thioploca, a chemolithotrophic sulphur-oxidizing genus of the proteobacteria family Thiotrichaceae, lives as filaments inside polysaccharide sheaths up to $500\,\mu m$ in diameter and up to $20\,cm$ long[19]. *Thioploca* accumulates nitrate intracellularly to concentrations four orders of magnitude higher than bottom-water nitrate[20]. Filaments glide vertically through the sediments within the sheaths, coupling sulphide oxidation at depth to dissimilatory nitrate reduction to ammonium (DNRA): $NO_3^- + H_2S + H_2O \rightarrow SO_4^{2-} + NH_4^+$ (refs 20, 21). *Thioploca*'s metabolism also results in the production of nitrite via the
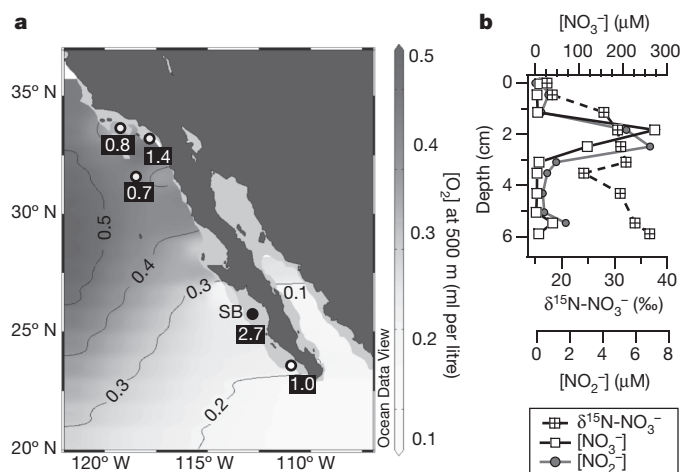

Figure 1 | Location and geochemistry of the Soledad basin sediments.
a, Map of Soledad basin location (SB). Black-boxed numbers are N_2 fluxes in mmol N $m^{-2}\,d^{-1}$; white circles, data from ref. 17; black circle, data from this work. Numbers on the contour lines are water-column $[O_2]$ at 500 m depth (based on data from World Ocean Atlas 2009; http://www.nodc.noaa.gov/OC5/WOA09/pr_woa09.html). The map was generated using Ocean Data View (http://odv.awi.de/). **b**, Nitrate and nitrite in pore waters extracted by whole-core squeezer[17] and $\delta^{15}N\text{-}NO_3^-$ (versus atmospheric N_2) determined as described in Supplementary Information section 1.4. Peaks in $[NO_3^-]$ and $[NO_2^-]$ result from bursting of *Thioploca* cells by squeezing the sediments.

[1]Department of Geology, Pomona College, Claremont, California 91711, USA. [2]Department of Earth Sciences, University of Southern California, 3651 Trousdale Parkway, Los Angeles, California 90089, USA. [3]Department of Microbiology and Molecular Genetics, University of California, Davis, 1 Shields Avenue, Davis, California 95616, USA. [4]Nordic Center for Earth Evolution, Department of Biology, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark. [5]Department of Biology, Pomona College, Claremont, California 91711, USA. [6]Department of Marine Sciences, University of Connecticut, Groton, Connecticut 06340, USA. [7]Department of Geosciences, Guyot Hall, Princeton University, Princeton, New Jersey 08544, USA. [8]Department of Geology, University of Cincinnati, 605 Geology-Physics Building, Cincinnati, Ohio 45221, USA.

intracellular deposition of granular sulphur (S^0): $NO_3^- + H_2S \rightarrow NO_2^- + S^0 + H_2O$ (refs 19, 21). Ammonium and nitrite generated by *Thioploca* are the required metabolic substrates for anammox bacteria, which obtain energy by oxidizing ammonium with nitrite to N_2: $NH_4^+ + NO_2^- \rightarrow N_2 + 2H_2O$ (ref. 14). Thus, *Thioploca* may be a desirable biogeochemical 'host' for anammox bacteria, but an association between the two organisms has not been demonstrated.

Here we present microbiological, molecular and isotopic evidence for the existence of a *Thioploca*–anammox consortium and show that N-based co-metabolism of the consortium drives a substantial fraction of fixed N loss in anoxic sediments.

The study was conducted in the eastern tropical North Pacific Ocean, within Soledad basin, a 544-m-deep basin separated from the open ocean by a sill at 250 m (Fig. 1a). Restricted circulation in the basin, combined with high C_{org} export from surface waters fuelled by coastal upwelling, results in an anoxic water column below the sill[22]. A large loss of fixed N was documented in the Soledad basin sediments in the form of an N_2 efflux of $2.7 \pm 0.4\,\mathrm{mmol\,N\,m^{-2}\,d^{-1}}$ across the sediment–water interface, along with a concurrent benthic NH_4^+ efflux of similar magnitude, $2.7 \pm 0.5\,\mathrm{mmol\,N\,m^{-2}\,d^{-1}}$ (ref. 17). An abundant

community of *Thioploca*[16,17,23] of about $(5 \pm 2) \times 10^4$ sheaths per square metre colonizes the rapidly accumulating (around $1–2\,\mathrm{mm\,yr^{-1}}$) organic-rich sediments (Supplementary Table 1A). A large subsurface pool of nitrate and nitrite, indicative of intracellular transport by *Thioploca* (Fig. 1b)[17], and an increased $^{15}N/^{14}N$ ratio in pore-water NH_4^+ (with a $\delta^{15}N$-NH_4^+ of up to 20‰ versus atmospheric N_2), suggesting anammox activity, were found in sediments of this basin (Supplementary Fig. 1a and Supplementary Table 1)[16].

Sediments and individual sheaths with filaments of *Thioploca* sp. were collected at 1 cm depth intervals for microbiological and molecular study (see Methods). Visually intact, vertically oriented, white *Thioploca* sheaths were present throughout the upper 6–8 cm. Below this depth, increasingly disintegrated sheaths were found, coincident with a sharp increase in pore-water H_2S (Supplementary Fig. 1a). Phylogenetic analysis of 16S ribosomal RNA gene sequences (Supplementary Information section 3.4) showed that *Thioploca* from the Soledad basin is probably *Thioploca araucae*, recently re-classified as *Candidatus Marithioploca araucae* (see Methods).

4′,6-diamidino-2-phenylindole (DAPI) staining of *Thioploca* sp. sheaths collected live from the 4–6 cm depth interval revealed a dense
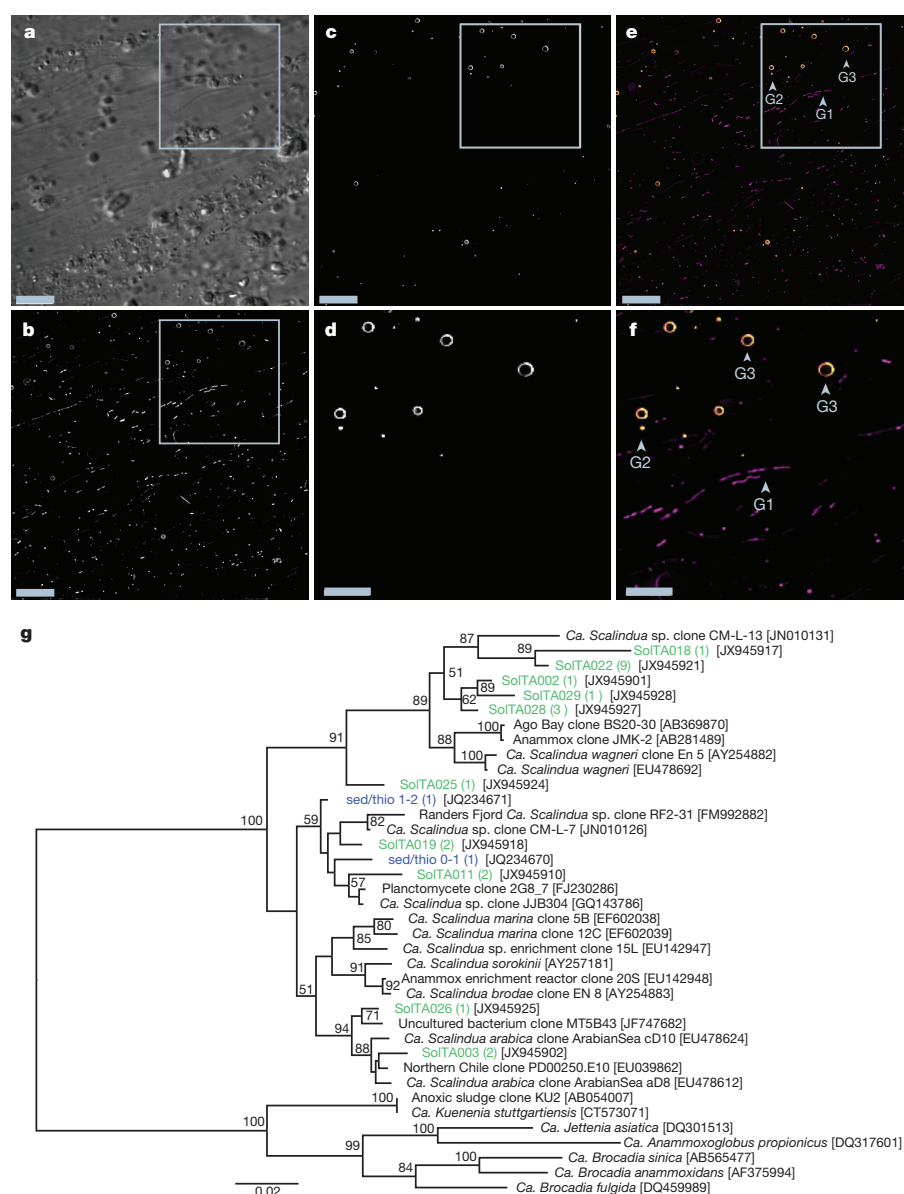
Figure 2 | Anammox bacteria found in close spatial association with *Thioploca* sheaths collected in the sediments of Soledad basin. **a**, Differential interference contrast image of a *Thioploca* sp. sheath, collected from the Soledad sediments in the 4–6-cm interval from multicore MC10. **b**, DAPI-stained *Thioploca* sp. sheath, highlighting the diverse morphologies of bacterial cells found within the *Thioploca* sp. sheath; the square defines the area shown in panels **a**, **c** and **e**). **c**, The AMX820 rDNA probe hybridized on the surface of *Thioploca* sp. sheaths with cells in a doughnut-shaped pattern, typical of anammox bacteria. **d**, Zoomed-in region of the white square in **c**, with more detailed view of doughnut-shaped cells, typical of anammox bacteria. **e**, Image overlay of **b** and **c** false-coloured with DAPI in magenta and the AMX820 rDNA probe in yellow; G1–G3 indicate the different bacterial morphologies on the surface of the *Thioploca* sp. sheath (see text). **f**, Zoomed-in region of the square in **e**, showing a close-up of the different bacterial morphologies (G1–G3). Images were collected using a Leica DMI6000 B inverted fluorescence microscope with differential interference contrast. Serial sections were acquired at 0.2-μm intervals; vertical data stacks were deconvolved using the Huygens deconvolution software (http://www.svi.nl/HuygensSoftware) and two-dimensional projections were created from the three-dimensional data sets using ImageJ. Scale bars in **c** and **e** are 10 μm, scale bars in **d** and **f** are 5 μm. **g**, Maximum-likelihood 16S rRNA gene phylogeny of anammox bacteria in the Soledad basin. Bootstrap values based on 1,000 replicates (over 50%) are indicated at each node. Phylotypes identified in Soledad basin sediments are in bold with numbers in parentheses indicating the number of clones detected in each phylotype. The *Candidatus Scalindua* DNA samples labelled 'SolTA' (green text) were obtained from *Thioploca* sheaths directly; the samples labelled 'sed/thio' (blue text) were from DNA extracted from sediments surrounding *Thioploca* sheaths.

population of microorganisms covering the sheaths (Fig. 2a–f). Morphologically, these cells could be broadly classified into three groups: two types of filamentous bacteria, 3–5 µm and 10–20 µm long (G1), cocci less than 0.5 µm in diameter (G2) and cells 1–2 µm in diameter DAPI-stained in a doughnut-shaped pattern (G3) (Fig. 2e, f).

An Alexa488 fluorescently labelled DNA probe (complementary to 16S rRNA), AMX820, that targets three known anammox genera (*Candidatus Brocadia, Candidatus Kuenenia* and *Candidatus Scalindua*) was applied to *Thioploca* sheaths (see Methods). The probe hybridized with G3-type cells in a doughnut-like pattern, typical of anammox bacteria because of a large intracellular anammoxosome devoid of genetic material[10] (G3, Fig. 2c–f). Fluorescent *in situ* hybridization (FISH) with the anammox-specific probe and the doughnut-shaped ribosomal and DNA-targeted staining pattern tentatively identified the G3 cells (Fig. 2a–f) as *Candidatus Scalindua*, the only marine anammox genus known to hybridize with the AMX820 probe (see Methods). One other phylum of bacteria, the *Poribacteria*, exhibits a similar doughnut-shaped morphology, but these are primarily found in association with sponges or in surrounding water, mostly in the presence of dissolved O₂ (ref. 24). However, in view of the two to three mismatches between known *Candidatus Scalindua* sequences and the AMX820 probe, FISH-based identification of G3 cells as an anammox organism remains provisional and must be considered only in the context of the other evidence presented below (see Methods and Supplementary Information section 3.5 for further details on the limitations of our FISH results). Some of the cocci (G2, Fig. 2e and f) also hybridized with the AMX820 probe, although their smaller size and the limit of resolution in fluorescence microscopy (about 250 nm) make it difficult to identify the shape of the staining pattern, so that their identification as anammox bacteria is less certain. Cells hybridizing with the AMX820 probe in the doughnut-shaped pattern were found in sediments surrounding *Thioploca* sheaths throughout the 0–8 cm depth interval (Supplementary Fig. 1b).

The presence of the anammox organisms was verified by the phylogenetic analysis of DNA material extracted from *Thioploca* sheaths and sediment samples (Fig. 2g). Amplified with anammox-specific primers 16S rRNA gene sequences placed the *Thioploca*-associated organisms within the anammox genus *Candidatus Scalindua* (see Methods and Supplementary Information section 3.4). The majority of *Thioploca*-associated anammox operational taxonomic groups (OTU) grouped with *Scalindua wagneri* and related taxa (Fig. 2g). However, two OTUs grouped with *Scalindua arabica*, and the remaining four (including two OTUs of sediment-extracted DNA) formed another group that does not seem to be closely related to previously characterized clades of *Scalindua* species. Amplification from the DNA material of *hzoA/hzoB* genes that encode one of the enzymes central to anammox metabolism placed the Soledad clones within the *Candidatus Scalindua* genus as well (see Supplementary Information section 1.2 for details and the Methods for GenBank accession numbers).

Anammox metabolic activity was documented by quantifying $^{29}N_2$ production in strictly anaerobic shipboard incubation experiments on $^{15}NH_4^+$-amended sediment slurries containing a natural population of *Thioploca* species (see Methods[11]). Of the three sediment intervals examined, 2–4 cm, 6–8 cm and 16–18 cm, the upper two showed $^{29}N_2$ production of 0.177 ± 0.014 and 0.042 ± 0.014 µmol N_2 per litre per hour respectively (Supplementary Table 1B, Supplementary Fig. 2). Given that neither nitrate nor nitrite was added, the anammox activity must have relied entirely on nitrite or nitrate available from *Thioploca*. Assuming a linear decrease in anammox rates with depth, the $^{29}N_2$ production integrated over the 0–8 cm interval amounted to N_2 flux of 0.53 ± 0.05 mmol N m^{-2} d^{-1}, or about 20% of the total measured N_2 efflux of 2.7 mmol N m^{-2} d^{-1} (ref. 17). $^{29}N_2$ production ceased after the first sampling time point (5 h of incubation; Supplementary Fig. 2), though neither nitrite nor ammonium was fully consumed. Therefore, anammox may have been inhibited by sulphide accumulating in the

incubations, potentially leading to underestimated incubation-based rates (Supplementary Information section 1.3).

Incubation-based $^{29}N_2$ production confirmed anammox activity, but the experimental rates may have been affected by sediment manipulation. Thus, we derived additional constraints on N_2 production by *Thioploca*–anammox consortia using a steady-state box model of the flux and isotope mass balances of nitrate uptake coupled to N_2 and NH_4^+ efflux (Fig. 3, Supplementary Table 2, and Supplementary Information section 1.4). In the model, the total measured benthic ammonium efflux (J_{totNH_4}, Table 1) is comprised of two NH_4^+ sources, organic nitrogen (N_{org}) degradation (J_{orgNH_4}) and DNRA ($J_{D-NH_4} = J_{totNH_4} - J_{orgNH_4}$) (Fig. 3a). The J_{orgNH_4} is determined by balancing

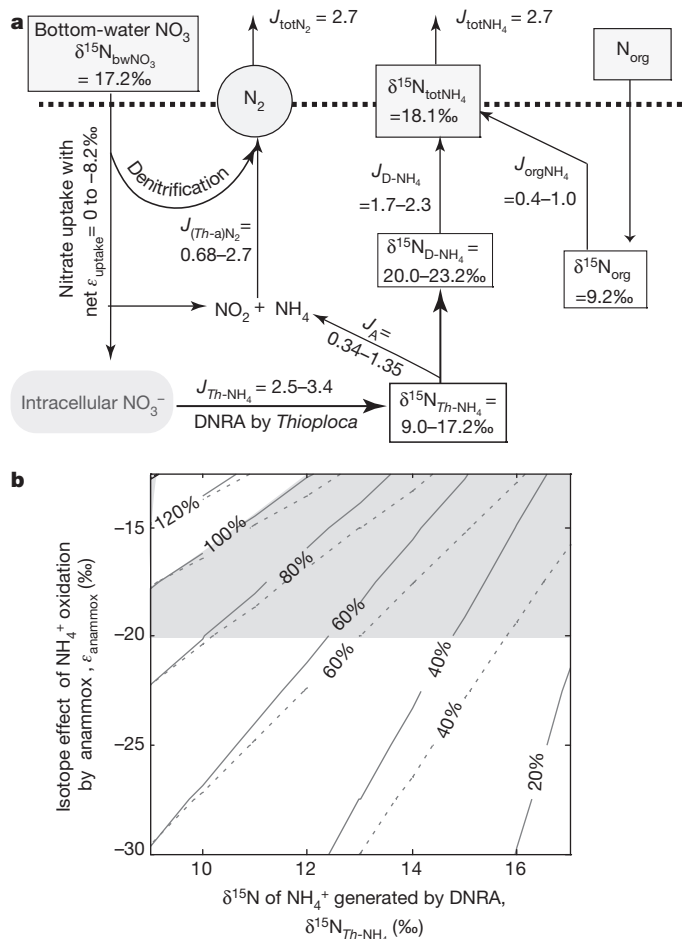**Figure 3 | N cycling by the *Thioploca*–anammox consortium and associated changes in δ¹⁵N of nitrate and ammonium.** **a**, Box model of the input and output N fluxes and δ¹⁵N (versus atmospheric N_2) of fixed N pools and fluxes measured or calculated as shown in the main text and in Supplementary Information section 2.2. The inputs include NO_3^-, either diffusing or transported into the sediments by *Thioploca*, and sedimentary N_{org} generating J_{orgNH_4} through decomposition. NO_3^- is converted to N_2 via denitrification, or to NO_2^- or NH_4^+ via DNRA by *Thioploca* (J_{Th-NH_4} flux). A fraction of J_{Th-NH_4} flux is oxidized by anammox with NO_2^-, generating the N_2 flux (the $J_{(Th-a)N_2}$ flux). The DNRA-generated NH_4^+ flux that remains after anammox oxidation (J_{D-NH_4}), together with the J_{orgNH_4} flux, comprises the benthic NH_4^+ efflux (J_{totNH_4}). The ranges of values shown correspond to the end-member values of C_{e^-}, $\varepsilon_{anammox}$ and ε_{uptake}, as discussed in the main text and shown in **b**. **b**, Calculated fractions of N_2 generated by *Thioploca*–anammox consortia, $F_{(Th-a)} = J_{(Th-a)N_2}/J_{totN_2}$ (shown as percentages). The dashed lines are $F_{(Th-a)}$ values calculated for the case of marine algal C_{org} as C source; solid lines are $F_{(Th-a)}$ values for CH_4 as C source. The $F_{(Th-a)}$ values are obtained for a δ¹⁵N$_{T-NH4}$ range of 9.0–17.2‰ and for an anammox isotope effect, $\varepsilon_{anammox}$, between −30‰ and −12‰. The $F_{(Th-a)}$ values obtained for the sub-range of $\varepsilon_{anammox}$ of −12‰ to −20‰ suggested in ref. 18 are shown within the grey-shaded area, excluding a subset of values greater than 100% in the upper left corner.

the number of electrons (e^-) transferred to nitrate in N_2 production via denitrification/anammox $(5e^-)$ and in NH_4^+ production via DNRA $(8e^-)$ to the number of electrons available from sedimentary C_{org} oxidation (see Supplementary Information section 2.1 for details). The benthic efflux of NH_4^+ and N_2 is probably the dominant sink of electrons generated from the C_{org} oxidation because efflux of other reduced compounds (methane, sulphide and Fe^{2+}/Mn^{2+}) is negligible in Soledad sediments[25]. Thus, the balance of electrons transferred from C_{org} (ultimate e^- donor) to NO_3^- (ultimate e^- acceptor) is:

$$\langle J_{totNH_4} - J_{orgNH_4} \rangle (8e^-) + \langle J_{totN_2} \rangle (5e^-) = J_{orgNH_4}(R_{C/N})C_{e^-} \quad (1)$$

where J_{totN_2} (total measured N_2 flux) combines the N_2 fluxes from denitrification and anammox. J_{orgNH_4} (NH_4^+ generated through N_{org} decomposition) is stoichiometrically linked to C_{org} oxidation through $R_{C/N}$, the C/N ratio of decomposing organic matter. C_{e^-} is the number of electrons available from C_{org} oxidation, determined by the C_{org} oxidation state. In Soledad sediments, a large fraction of H_2S is produced by methane oxidation with sulphate[25], so that methane (C^{4-}, $C_{e^-} = 8$) and marine algal C_{org} (C^0, $C_{e^-} = 4$) are the two likely end-member C electron sources for nitrate reduction to N_2 or NH_4^+. The calculated ranges for J_{orgNH_4} (0.4–1.0 mmol N m^{-2} d^{-1}) and J_{D-NH_4} (1.7–2.3 mmol N m^{-2} d^{-1}) reflect the range between these end-members and a $R_{C/N}$ range of 6.7 to 9 (Supplementary Information section 2.1).

With J_{orgNH_4} and J_{D-NH_4} fluxes calculated, the $\delta^{15}N$ of DNRA-derived NH_4^+ ($\delta^{15}N_{D-NH_4}$) is determined from the isotopic mass balance, $J_{D-NH_4} \times \delta^{15}N_{D-NH_4} + J_{orgNH_4} \times \delta^{15}N_{org} = J_{totNH_4} \times \delta^{15}N_{totNH_4}$, where the $\delta^{15}N_{org}$ and $\delta^{15}N_{totNH_4}$ are known (Table 1 and Supplementary Information section 2.2). Assuming no isotopic fractionation for N_{org} decomposition in these sediments[16], the $\delta^{15}N$ of the DNRA-derived NH_4^+ pool ($\delta^{15}N_{D-NH_4}$) is calculated as 20.0–23.2‰, the range reflecting the two C source end-members.

At steady state, the $\delta^{15}N$ of $Thioploca$-generated NH_4^+ ($\delta^{15}N_{Th-NH_4}$) (Table 1 and Fig. 3a) should reflect bottom-water nitrate, $\delta^{15}N_{bwNO_3}$, modified by an isotope effect associated with nitrate uptake by $Thioploca$ (ε_{uptake}), so that $\delta^{15}N_{Th-NH_4} = \delta^{15}N_{bwNO_3} + \varepsilon_{uptake}$. From measured $\delta^{15}N_{bwNO_3}$ of 17.2‰ and an ε_{uptake} between 0‰ and −8.2‰ (see Supplementary information sections 1.4 and 2.2), the $\delta^{15}N_{Th-NH_4}$ is 9.0–17.2‰ (for the range of ε_{uptake}). These $\delta^{15}N_{Th-NH_4}$ values are

3–14‰ lower than the 20.0–23.2‰ range of the $\delta^{15}N_{D-NH_4}$, which is consistent with partial oxidation of J_{Th-NH_4} by the anammox cells observed in close association with the $Thioploca$ filaments (Fig. 2).

In Supplementary Information section 2.2, we show that the fraction of $Thioploca$-generated ammonium that is oxidized by anammox bacteria ($f = J_A/J_{Th-NH_4}$, where $J_A = J_{Th-NH_4} - J_{D-NH_4}$ is NH_4^+ oxidation by anammox) can be determined as:

$$f = \frac{\delta^{15}N_{Th-NH_4} - \delta^{15}N_{D-NH_4}}{\delta^{15}N_{totNH_4} + \varepsilon_{anammox} - \delta^{15}N_{D-NH_4}} \quad (2)$$

where $\varepsilon_{anammox}$ is the isotope effect of anammox with respect to NH_4^+ consumption. The value of f is estimated for the range of $\varepsilon_{anammox}$ between −12‰ and −30‰ (see Supplementary Information section 2.2 for discussion of the chosen range). The ammonium oxidation flux, J_A, is obtained from $J_A = f \times J_{Th-NH_4}$ (Table 1 and Supplementary Information section 2.2). Because J_A represents only half of the N_2 production by anammox (the second N atom deriving from nitrite), the fractional contribution of $Thioploca$–anammox-generated N_2 ($J_{(Th-a)N_2}$) to the total measured benthic N_2 efflux (J_{totN_2}) can be calculated as $F_{(Th-a)} = 2 \times J_A/J_{totN_2}$. For the C^{4-} (methane), the median $J_{(Th-a)N_2}$ was calculated to be 1.5 ± 0.6 ($\pm$s.d.) mmol N m^{-2} d^{-1} ($F_{(Th-a)} = 55 \pm 23\%$, Fig. 3b, solid lines); for C^0 (C_{org}), the median of $J_{(Th-a)N_2}$ was 1.6 ± 0.5 ($\pm$s.d.) mmol N m^{-2} d^{-1} ($F_{(Th-a)} = 60 \pm 19\%$, Fig. 3b, dashed lines).

To summarize the model results, about $57 \pm 21\%$ of N_2 production is derived from anammox relying on $Thioploca$ for the supply of nitrite. Despite the potential artefacts of the sediment incubations and simplifying assumptions made in the model (for example, steady state), the model-based estimate of the $Thioploca$–anammox N_2 production agrees within a factor of two with the ^{29}N incubation-based anammox rates.

$Thioploca$ serves as a beneficial host for anammox bacteria by providing the metabolically required substrates (as summarized in Supplementary Fig. 7 and Supplementary Information). However, whether the described symbiosis represents mutualism, syntrophy or commensalism remains uncertain. At sulphidic boundaries, $Thioploca$ may need to rapidly convert toxic HS^- to elemental S^0, generating nitrite (Supplementary Fig. 6). Nitrite is a potentially toxic metabolite, so active nitrite removal by anammox would be advantageous for $Thioploca$, benefiting both bacteria (mutualism). If $Thioploca$ actually requires anammox to remove nitrite, this mutualism would be classified as a syntrophy. Alternatively, anammox may simply be using nitrite leaking from $Thioploca$ with no benefit to $Thioploca$ (commensalism).

Analysis of sedimentary N fluxes shows that the $Thioploca$–anammox consortium enhances the benthic loss of fixed N that is otherwise limited by the diffusive supply of solutes in anoxic sediments[15]. Fixed N losses via all forms of denitrification are ultimately driven by the supply of sinking C_{org}, the fraction of primary production exported from the surface ocean. Fixed N loss, in turn, may act as a negative feedback on primary production through N limitation of phytoplankton growth[9], thus stabilizing the biological C cycle.

The degree of coupling between the N and C cycles through the sedimentary component of this feedback can be diagnosed from the molar ratio of fixed N loss to total C_{org} oxidation, N_{loss}/C_{oxid}. Globally, N_{loss}/C_{oxid} initially increases with decreasing bottom water $[O_2]$ (arrow 1 in Fig. 4) and is then expected to decrease in sediments underlying suboxic/anoxic bottom waters owing to the diffusion limitation of the nitrate supply below the sediment-water interface[15,26,27] (arrow 2 in Fig. 4). In the sediments of Soledad basin, total depth-integrated C_{org} oxidation is 3.3 ± 1.0 mmol C m^{-2} d^{-1} (see Supplementary Information section 2.3). The resulting N_{loss}/C_{oxid} ratio of 0.82, higher than in other sedimentary environments lacking the $Thioploca$–anammox assemblages (Fig. 4 and refs 17, 18, 27–29), suggests that the $Thioploca$-anammox symbiosis acts to tighten the sedimentary C_{org}-supply/N-loss feedback. $Thioploca$ species occurrences have been documented in anoxic sediments globally[19,30]. If widespread, the $Thioploca$–anammox symbiosis may represent an important sedimentary
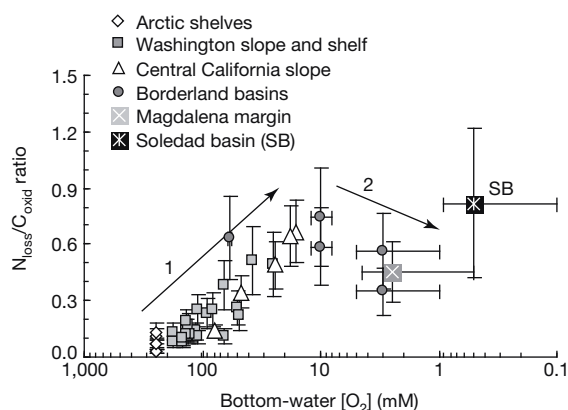
Table 1 | Measured and calculated benthic N fluxes

Measured benthic N fluxes and $\delta^{15}N$ values		
J_{totNH_4}	Total benthic NH_4^+ flux	2.7 ± 0.5 mmol N m^{-2} d^{-1}
J_{totN_2}	Total benthic N_2 flux	2.7 ± 0.4 mmol N m^{-2} d^{-1}
$\delta^{15}N_{org}$	$\delta^{15}N$ of sedimentary organic N	9.2 ± 0.2‰
$\delta^{15}N_{bwNO_3}$	$\delta^{15}N$ of bottom water NO_3^-	17.2 ± 0.1‰
$\delta^{15}N_{totNH_4}$	$\delta^{15}N$ of the total ammonium efflux	18.1 ± 1.1‰
ε_{uptake}	Net isotopic effect of sedimentary nitrate uptake	-4.1 ± 1.2‰
Calculated benthic N fluxes and $\delta^{15}N$ values		
J_{orgNH_4}	NH_4^+ efflux from N_{org} degradation	0.4–1.0 mmol N m^{-2} d^{-1}*
J_{D-NH_4}	NH_4^+ efflux from DNRA (after anammox)	1.7–2.3 mmol N m^{-2} d^{-1}*
J_{Th-NH_4}	NH_4^+ flux generated by $Thioploca$	2.3–3.0 mmol N m^{-2} d^{-1}
$J_A = J_{Th-NH_4} - J_{D-NH_4}$	NH_4^+ oxidized by anammox	0.34–1.35 mmol N m^{-2} d^{-1}*
$\delta^{15}N_{D-NH_4}$	$\delta^{15}N$ of NH_4^+ from DNRA (after anammox)	20.0–23.2‰*
$\delta^{15}N_{Th-NH_4}$	$\delta^{15}N$ of NH_4^+ generated by $Thioploca$	9.0–17.2‰

* The range refers to variable oxidation states of C, C^{4-} (CH_4) and C^0 (C_{org}). Major nitrogen fluxes across the sediment–water interface and isotopic compositions ($\delta^{15}N$ versus air N_2) of major fixed N pools and fluxes were measured as described in Supplementary Information section 1.4. J_{orgNH_4} and J_{D-NH_4} were calculated using Supplementary equations (1) and (2). $\delta^{15}N_{D-NH_4}$ was calculated using Supplementary equation (3). J_{totNH_4} was calculated using Supplementary equations (4) to (8). See Supplementary Information sections 2.1–2.2.

Figure 4 | Relationship between benthic N_{loss}/C_{oxid} ratio and bottom-water $[O_2]$. Values calculated from published rates of sedimentary C_{org} oxidation rates (C_{oxid}) and fixed N_{loss} values are given for comparison[17,18,27–29]. Error bars represent one standard error based on reported uncertainties in $[O_2]$ measurements and flux determinations.

sink for fixed N and may become increasingly important in a warming, less oxygenated ocean.

METHODS SUMMARY

Multicores (SOL MC8 and SOL MC10) were collected in Soledad basin in 2009 during a cruise on the RV *New Horizon*, at a depth of 544 m (station coordinates: 25° 12.49′ N/112° 42.27′ W and 25° 12.33′ N/112° 42.09′ W).

Cores were sectioned in 1-cm intervals (upper 10 cm) or 3-cm intervals (below 10 cm to the bottom of the core). *Thioploca* sheaths carefully extracted from the sediment and sediment samples used for FISH analysis were fixed in 4% para-formaldehyde, embedded in polyacrylamide, and attached to coverslips. Coverslips were pre-treated with a 10 mg ml^{-1} solution of lysozyme at room temperature and subsequently at 37 °C for 20 min and incubated at 46 °C overnight with hybridization buffer containing 10 ng μl^{-1} AMX 820 probe (see Methods). After 18 h, coverslips were incubated with wash buffer at 46 °C for 20 min twice and mounted in Prolong Gold Antifade Reagent with DAPI (also see Methods).

Partial 16S small subunit rRNA genes were amplified using anammox-specific primers and the hydrazine oxido-reductase genes *hzoA/hzoB* were amplified with nested primer pairs (hzoAB1F, hzoAB1R) and (hzoAB4F, hzoAB4R), or with (hzo888f, hzo1101r) (see Methods, Supplementary Information sections 1.2 and 3.4 and Supplementary Table 4).

Three cores were incubated for 24 h at *in situ* temperature, and overlying water was sampled at regular time intervals for determination of $\delta^{15}NH_4^+$ and $\delta^{15}NO_3^-$ of the corresponding fluxes (details in Supplementary Information sections 1.4 and 3.1–3.3).

For determination of anammox activity, sediment cores were sectioned in 2-cm intervals under anaerobic conditions, aliquots were mixed into a slurry with low-nutrient surface sea water, and after addition of $^{15}NH_4$ label (no $^{15}NO_3/NO_2$ was added), incubated for a total of 22 h. Total anammox activity was calculated based on the excess production of $^{29}N_2$ accounting for isotope dilution[11] (see Methods).

Full Methods and any associated references are available in the online version of the paper.

1. Biebl, H. & Pfennig, N. Growth yields of green sulfur bacteria in mixed cultures with sulfur and sulfate reducing bacteria. *Arch. Microbiol.* **117,** 9–16 (1978).
2. Schink, B. & Stams, A. J. M. In *Prokaryotes: A Handbook on the Biology of Bacteria* Vol. 2, 3rd edn *Ecophysiology and Biochemistry* (eds Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K. H. & Stackebrandt, E.) 309–335 (Springer, 2006).
3. Wang, G. Z., Spivack, A. J. & D'Hondt, S. Gibbs energies of reaction and microbial mutualism in anaerobic deep subseafloor sediments of ODP Site 1226. *Geochim. Cosmochim. Acta* **74,** 3938–3947 (2010).
4. Valentine, D. L. Biogeochemistry and microbial ecology of methane oxidation in anoxic environments: a review. *Antonie van Leeuwenhoek* **81,** 271–282 (2002).
5. Hoehler, T. M., Alperin, M. J., Albert, D. B. & Martens, C. S. Field and laboratory studies of methane oxidation in an anoxic marine sediment — evidence for a methanogen-sulfate reducer consortium. *Glob. Biogeochem. Cycles* **8,** 451–463 (1994).
6. Orphan, V. J., House, C. H., Hinrichs, K. U., McKeegan, K. D. & DeLong, E. F. Methane-consuming archaea revealed by directly coupled isotopic and phylogenetic analysis. *Science* **293,** 484–487 (2001).
7. Boetius, A. *et al.* A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature* **407,** 623–626 (2000).
8. Orphan, V. J. Getting cozy: hidden microbial interactions in nature. *Environ. Microbiol. Rep.* **3,** 16–18 (2011).
9. Gruber, N. in *The Ocean Carbon Cycle and Climate* (eds Follows, M. & Oguz, T.) 97–148 (NATO ASI Series, Kluwer Academic, 2004).
10. Kuenen, J. G. Anammox bacteria: from discovery to application. *Nature Rev. Microbiol.* **6,** 320–326 (2008).
11. Thamdrup, B. & Dalsgaard, T. Production of N_2 through anaerobic ammonium oxidation coupled to nitrate reduction in marine sediments. *Appl. Environ. Microbiol.* **68,** 1312–1318 (2002).
12. Carpenter, E. J. *et al.* Extensive bloom of a N_2-fixing diatom/cyanobacterial association in the tropical Atlantic Ocean. *Mar. Ecol. Prog. Ser.* **185,** 273–283 (1999).
13. Foster, R. A. & Zehr, J. P. Characterization of diatom-cyanobacteria symbioses on the basis of nifH, hetR and 16S rRNA sequences. *Environ. Microbiol.* **8,** 1913–1925 (2006).
14. Bernhard, J. M., Edgcomb, V. P., Casciotti, K. L., McIlvin, M. R. & Beaudoin, D. J. Denitrification likely catalyzed by endobionts in an allogromiid foraminifer. *ISME J.* **6,** 951–960 (2012).
15. Middelburg, J. J., Soetaert, K., Herman, P. M. J. & Heip, C. H. R. Denitrification in marine sediments. *Glob. Biogeochem. Cycles* **10,** 661–673 (1996).
16. Prokopenko, M. *et al.* Nitrogen cycling in the sediments of Santa Barbara basin and Eastern Subtropical North Pacific: nitrogen isotopes, diagenesis and possible chemosymbiosis between two lithotrophs *Thioploca* and anammox: riding on a glider. *Earth Planet. Sci. Lett.* **242,** 186–204 (2006).
17. Chong, L., Prokopenko, M. G., Berelson, W. M., Townsend-Small, A. & McManus, J. Nitrogen cycling within suboxic and anoxic marine sediments from the continental margin of Western North America. *Mar. Chem.* **128–129,** 13–25 (2012).
18. Prokopenko, M. *et al.* Denitrification in anoxic sediments supported by biological nitrate transport. *Geochim. Cosmochim. Acta* **75,** 7180–7199 (2011).
19. Jørgensen, B. B. & Gallardo, V. A. *Thioploca* spp: filamentous sulfur bacteria with nitrate vacuoles. *FEMS Microbiol. Ecol.* **28,** 301–313 (1999).
20. Fossing, H. *et al.* Concentration and transport of nitrate by mat-forming sulphur bacterium *Thioploca. Nature* **374,** 713–715 (1995).
21. Otte, S. *et al.* Nitrogen, carbon, and sulfur metabolism in natural *Thioploca* samples. *Appl. Environ. Microbiol.* **65,** 3148–3157 (1999).
22. van Geen, A., Smethie, W. M., Horneman, A. & Lee, H. Sensitivity of the North Pacific oxygen minimum zone to changes in ocean circulation: a simple model calibrated by chlorofluorocarbons. *J. Geophys. Res.* **111,** C10004 (2006).
23. Bernhard, J. M. & Buck, K. R. in *Sulfur Biogeochemistry — Past and Present* GSA Spec. Pap. 379 (eds Amend, J. P., Edwards, K. J. & Lyons, T. W.) 35–47 (Geological Society of America, 2004).
24. Webster, N. S. *et al.* Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. *Environ. Microbiol.* **12,** 2070–2082 (2010).
25. Berelson, W. *et al.* Anaerobic diagenesis of silica and carbon in continental margin sediments: discrete zones of TCO_2 production. *Geochim. Cosmochim. Acta* **69,** 4611–4629 (2005).
26. Christensen, J. P., Murray, J. W., Devol, A. H. & Codispoti, L. A. Denitrification in continental shelf sediments has major impact on the oceanic nitrogen budget. *Glob. Biogeochem. Cycles* **1,** 97–116 (1987).
27. Hartnett, H. E. & Devol, A. H. Role of a strong oxygen-deficient zone in the preservation and degradation of organic matter: a carbon budget for the continental margins of northwest Mexico and Washington State. *Geochim. Cosmochim. Acta* **67,** 247–264 (2003).
28. Devol, A. H., Codispoti, L. A. & Christensen, J. P. Summer and winter denitrification rates in western Arctic shelf sediments. *Cont. Shelf Res.* **17,** 1029–1033 (1997).
29. Berelson, W. M. *et al.* Biogenic matter diagenesis on the sea floor: a comparison between two continental margin transects. *J. Mar. Res.* **54,** 731–762 (1996).
30. Høgslund, S. *et al.* Physiology and behaviour of marine *Thioploca. ISME J.* **3,** 647–657 (2009).

METHODS

Collection of *Thioploca sp.* and sediments for DNA extraction and FISH. Multicores (SOL MC8 and SOL MC10) were collected at a depth of 544 m, at 25° 12.49′ N and 112° 42.27′ W and 25° 12.33′ N and 112° 42.09′ W respectively. Samples of *Thioploca* sheaths with filaments and sediment aliquots were collected in 1-cm intervals down to 10 cm depth and then every 3 cm down to the bottom of the core. *Thioploca* sp. samples were washed twice with overlying sea water and spun down at 900g (where g is the acceleration due to gravity) for 5 min at room temperature (about 27 °C) to remove sediment from the *Thioploca* sp. sheaths and filaments. *Thioploca* samples and sediments collected in 1.5-ml tubes were stored at −80 °C for further DNA extraction and analysis (see Supplementary Information section 3.4). Both types of samples (*Thioploca* sp. sheaths and sediments) to be analysed with FISH were resuspended in 400 µl of 1× Hepes buffered saline (HBS), and fixed with paraformaldehyde (with a final concentration of 4%) for 30 min at room temperature while rotating. Samples were spun at 900g for 5 min at room temperature, and, after removing the supernatant, washed in 1 ml of PEM (comprising 100 mM piperazine-N,N′-bis(2-ethanesulphonic acid) (PIPES) plus 1 mM ethylene glycol tetraacetic acid (EGTA) plus 0.1 mM MgSO₄) twice. Samples were re-suspended in PEM and stored at 4 °C.

Thioploca sp. determination. Genomic DNA was extracted from sediments using the MoBio PowerBioFilm DNA isolation kit. The 16S rRNA gene specific to *Thioploca* was amplified using a broad γ-Proteobacteria forward primer (109f2), and a reverse primer specific to *Thioploca* spp. (829-Thioploca)[31]. Amplification products were cloned using the TOPO TA Cloning kit for sequencing (Invitrogen). Twenty positive transformants were selected for sequencing. Sequences were checked for chimaeras using Mallard 1.02 (ref. 32) before phylogenetic analysis. Phylogenetic analysis of 16S rRNA gene sequences (Supplementary Information section 3.4) showed that *Thioploca* from the Soledad basin is probably *Thioploca araucae*[33], recently re-classified as *Candidatus Marithioploca araucae*[34].

FISH using the probe AMX820. Prior to FISH, fixed *Thioploca* sp. (from sample SOL MC10, 4–6 cm) and sediment splits (from sample SOL MC8, 4–5 cm) were attached to slides using activated coverslips crosslinked to polyacrylamide sheets[35]. *Thioploca* sheaths or sediment were embedded in thin sheets of polyacrylamide gel that were covalently attached to activated coverslips. Two pieces of *Thioploca* sp. were placed onto activated coverslips, 25 µl of polyacrylamide solution was pipetted over *Thioploca* sp., and the droplet was flattened using a large circular coverslip (number 1.5, 22-mm diameter). The sandwich was polymerized for 30 min, the circular glass coverslip removed with a razor blade, and the gel washed on a shaker with 50 mM HEPES pH = 8.5.

Twenty microlitres of a 10 mg ml⁻¹ solution of lysozyme (in autoclaved distilled water) were placed in the bottom of a culture dish plate. Embedded *Thioploca* sp. or sediment coverslips were placed face-down on the lysozyme solution at room temperature for 20 min, and subsequently placed at 37 °C for 20 min. Embedded *Thioploca* sp. sheaths and sediment coverslips were washed in a 1× phosphate buffered saline (PBS) solution twice, and hybridized with pre-warmed hybridization buffer containing 5 µl of 100 ng µl⁻¹ S-*-Amx-0820-a-A-22 probe labelled with fluorescent dye Alexa488 (AMX 820, sequence: 5′-AAAACCCCTCTACTT AGTGCCC-3′)[36], which targets the genera *Candidatus Brocadia*, *Candidatus Kuenenia* and *Candidatus Scalindua*[37], or no probe as negative control. The hybridization buffer contained distilled, autoclaved water, 0.9 M NaCl, 0.02 M Tris-HCl, pH = 8, and 0.01% SDS (sodium dodecyl sulphate) (0% formamide) and hybridization occurred overnight in a 46 °C water bath in a humidifying chamber.

Our *Candidatus Scalindua* sequences showed 10–12% mismatches when compared to both *Candidatus Brocadia* and *Candidatus Kuenenia* 16S rRNA gene sequences of homologous, conserved positions using the ARB database[38] (accession numbers of sequences used for comparison are: AF375995, CT573071, EU478693, PQ459989, AF375994). Furthermore, the AMX820 probe was found to have two mismatches with *Scailindua arabica*, three mismatches with *S. waneri*, and between two and three mismatches with *S. brodae* and *S. marina*. Thus, hybridization conditions were optimized by varying formamide concentration, salt concentration (wash buffer), and temperature with respect to the potential mismatches between the AMX820 probe and *Candidatus Scalindua* species.

The following morning, embedded *Thioploca* filament or sediment coverslips were washed with wash buffer containing the following: distilled, autoclaved water, 0.9 M NaCl, 0.02 M Tris-HCl, pH = 8, and 0.01% SDS. Coverslips were washed for 20 min at 46 °C twice and then mounted in 30 µl Prolong Gold Antifade reagent with DAPI. Slides were dried and stored in a cool, dry place for at least 20 h before visualizing. Images were collected using a Leica DMI6000 B inverted fluorescence microscope with differential interference contrast. Serial sections were acquired at 0.2-µm intervals. Data stacks were deconvolved using the Huygens deconvolution

software and two-dimensional projections were created from the three-dimensional data sets using ImageJ.

Sample hybridization with the AMX820 probe and the 'no probe' negative controls were conducted on the same day and imaged simultaneously. 'No probe' controls revealed no autofluorescence of doughnut-shaped cells or background autofluorescence in the sample. At the request of a reviewer, two years after the main experiments were conducted, a 'nonsense' probe control, RPE0108Dasy, was hybridized with *Thioploca* sheaths. The 'nonsense' probe was labelled with the same amino-terminal Alexa488 fluorescent dye as was used in the original FISH experiment with the AMX820 probe. The hybridization conditions were identical to those used in the initial experiment. The 'nonsense' probe did not hybridize with *Thioploca* sheaths or bacterial cell walls (see Supplementary Fig. 8), indicating the absence of unspecific binding of the Alexa488 dye to any material in the sample (see Supplementary Information section 3.5 for further details on the limitations of the 'nonsense' probe experiment, including a discussion of the ageing of the samples). All images (including images of the 'nonsense' probe hybridization) were collected with the same fluorescence intensity, the same exposure time, and subject to the same deconvolution parameters using the Huygens deconvolution software.

Taking into account modifications in stringency of hybridization conditions and the overall 10–12% mismatches between the AMX 820 probe and the targeted sequences, identification of *Thioploca*-associated anammox based on rRNA-targeted FISH with the AMX 820 probe should be considered preliminary until it is confirmed with a 100% matching probe or a Clone-FISH procedure.

PCR amplification of anammox bacteria DNA. To further verify the presence of anammox bacteria, DNA material was extracted from sediment samples and *Thioploca* sheaths and amplified with polymerase chain reaction (PCR) (see Supplementary Information section 3.4 for details on DNA extraction). Partial 16S small subunit rRNA genes were amplified using anammox-specific primers, An7F[39] with AMX 820R[36], (2) S-P-Planc-0046-18F[36] with S-*-Amx-0368-a-A-18r[36] (Supplementary Table 4) with PCR products cloned and sequenced. Sequences with over 98.7% sequence identity formed 12 operational taxonomic groups that represented broad diversity within the marine genus of anammox bacteria, *Scalindua* (Fig. 2G in the main text and Supplementary Information section 1.2). In addition, the hydrazine oxido-reductase genes, *hzoA/hzoB*, encoding one of the enzymes central for anammox metabolism, were amplified with nested primer pairs hzoAB1F, hzoAB1R; and hzoAB4F, hzoAB4R[40] (Supplementary Fig. 1c). *Hzo*-gene and partial 16S rRNA gene sequences have been submitted to GenBank (accession numbers JQ234655–672, JX945900-903, JX945905, JX945907-908, JX945910, JX945913, JX945915, JX945917-919, JX945921-928, JX945930 and JX945932-963).

Sediment ¹⁵NH₄⁺ incubations for determining anammox activity. Following protocol described in ref. 11, sediments were collected by the multicorer onboard the ship and immediately placed in a cold room at *in situ* temperature (10 °C). The sediment core was sliced inside a polyethylene glove bag flushed twice with N₂ before closing. Slices collected from 2–4 cm, 6–8 cm and 16–18 cm depth intervals were kept inside the glove bag for ¹⁵NH₄⁺ amendments to measure anammox activity. The 2–4 cm and 6–8 cm intervals were chosen based on the observed live *Thioploca* sp. distribution, whereas the 16–18 cm interval was chosen as the lowest sediment depth into which *Thioploca* sp. visibly extended.

Homogenized sediments were subsampled for porosity and initial ammonium concentrations. Concentrated ¹⁵NH₄Cl solution (100 mmol per litre) was added to final ¹⁵NH₄⁺ fractions (F_N) between 0.42 and 0.62. Homogenized sediments were sampled for final ammonium concentrations 10 min after tracer addition to allow for sediment adsorption of NH₄⁺. Finally, sediments were distributed over ten 12-ml Exetainers (Labco limited, UK), each containing 9 ml deoxygenated artificial sea water (Red Sea salt solution of salinity 35.2 g kg⁻¹, or 35 practical salinity units).

Time series incubations lasted 22 h, during which two Exetainers were sacrificed every 5 h (Supplementary Fig. 2). Five millilitres of sediment slurry was removed and replaced by a helium headspace. Microbial activity was stopped by injecting 200 µl of a 50% (W/V) ZnCl₂ solution into the Exetainers. Two millilitres of sediment slurry was immediately injected into a He-flushed 12-ml Exetainer for mass spectrometric determination of ²⁸N₂, ²⁹N₂ and ³⁰N₂. The leftover slurry was filtered and kept frozen for [NH₄⁺] measurements and mass spectrometric determination of the labelled NH₄⁺ fraction in the sediment.

Ammonium concentrations were determined using the flow injection method with conductivity detection[41]. The fraction of ¹⁵NH₄⁺ in the NH₄⁺ pool was determined through hypobromite conversion of NH₄⁺ to N₂ (refs 42, 43). Total anammox activity was calculated based on the excess production of ²⁹N₂ and F_N (ref. 11).

31. Kojima, H., Teske, A. & Fukuki, M. Morphological and phylogenetic characterizations of freshwater *Thioploca* species from Lake Biwa, Japan, and Lake Constance, Germany. *Appl. Environ. Microbiol.* **69,** 390–398 (2003).

32. Ashelford, K. E., Chuzhanova, N. A., Fry, J. C., Jones, A. J. & Weightman, A. J. New screening software shows that most recent large 16S rRNA gene clone libraries contain chimeras. *Appl. Environ. Microbiol.* **72,** 5734–5741 (2006).

33. Teske, A., Ramsing, N. B., Kuver, J. & Fossing, H. Phylogeny of *Thioploca* and related filamentous sulfide-oxidizing bacteria. *Syst. Appl. Microbiol.* **18,** 517–526 (1995).

34. Salman, V. *et al.* A single-cell sequencing approach to the classification of large, vacuolated sulfur bacteria. *Syst. Appl. Microbiol.* **34,** 243–259 (2011).

35. Dembo, M. & Wang, Y. L. Stresses at the cell-to-substrate interface during locomotion of fibroblasts. *Biophys. J.* **76,** 2307–2316 (1999).

36. Schmid, M. C. *et al.* Biomarkers for in situ detection of anaerobic ammonium-oxidizing (anammox) bacteria. *Appl. Environ. Microbiol.* **71,** 1677–1684 (2005).

37. Jayakumar, A., O'Mullan, G. D., Naqvi, S. W. A. & Ward, B. B. Denitrifying bacterial community composition changes associated with stages of denitrification in oxygen minimum zones. *Microb. Ecol.* **58,** 350–362 (2009).

38. Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Res.* **32,** 1363–1371 (2004).

39. Penton, C. R., Devol, A. H. & Tiedje, J. M. Molecular evidence for the broad distribution of anaerobic ammonium-oxidizing bacteria in freshwater and marine sediments. *Appl. Environ. Microbiol.* **72,** 6829–6832 (2006).

40. Hirsch, M. D., Long, Z. T. & Song, B. Anammox bacterial diversity in various aquatic ecosystems based on the detection of hydrazine oxidase genes (hzoA/hzoB). *Microb. Ecol.* **61,** 264–276 (2011).

41. Hall, P. O. J. & Aller, R. C. Small-volume, flow injection analysis for CO_2 and NH_4^+ in marine and freshwaters. *Limnol. Oceanogr.* **37,** 1113–1119 (1992).

42. Risgaard-Petersen, N., Rysgaard, S. & Revsbech, N. P. Combined microdiffusion-hypobromite oxidation method for determining ^{15}N isotope in ammonium. *Soil Sci. Soc. Am. J.* **59,** 1077–1080 (1995).

43. Jensen, E. S. *et al.* Intensive nitrogen loss over the Omani shelf due to anammox coupled with dissimilatory nitrite reduction to ammonium. *ISME J.* **5,** 1660–1670 (2011).

LETTER

A new arboreal haramiyid shows the diversity of crown mammals in the Jurassic period

Xiaoting Zheng[1,2], Shundong Bi[3,4], Xiaoli Wang[1,2] & Jin Meng[3,5]

A major unsolved problem in mammalian evolution is the origin of Allotheria, including Multituberculata and Haramiyida[1–5]. Multituberculates are the most diverse and best known Mesozoic era mammals and ecologically resemble rodents, but haramiyids are known mainly from isolated teeth, hampering our search for their phylogenetic relationships. Here we report a new haramiyid from the Jurassic period of China, which is, to our knowledge the largest reported so far. It has a novel dentition, a mandible resembling advanced multituberculates and postcranial features adapted for arboreal life. Our phylogenetic analysis places Haramiyida within crown Mammalia, suggesting the origin of crown Mammalia in the Late Triassic period and diversification in the Jurassic, which contrasts other estimated divergence times of crown Mammalia[6–8]. The new haramiyid reveals additional mammalian features of the group, helps to identify other haramiyids represented by isolated teeth, and shows again that, regardless of various phylogenetic scenarios, a complex pattern of evolution involving many convergences and/or reversals existed in Mesozoic mammals.

Mammalia Linnaeus, 1758
Allotheria Marsh, 1880
Haramiyida Hahn, Sigogneau-Russell and Wouters, 1989
Arboroharamiyidae gen. nov.
Arboroharamiya jenkinsi gen. et sp. nov.

Etymology. *arbor* (Latin): tree; *Haramiya* (Arabic): trickster, petty thief; *jenkinsi*: in honour of Farish A. Jenkins Jr for his contribution to the study of Mesozoic mammals, haramiyids included.

Holotype. A partial skeleton with both mandibles associated with teeth and isolated upper teeth (STM33-9, Tianyu Museum of Nature, Shandong Province, China; Fig. 1 and Supplementary Figs 1–7).

Locality and horizon. The holotype is from the Middle–Late Jurassic Tiaojishan Formation in the town of Mutoudeng, Hebei Province, China, dated about 160 million years (Myr)[9] (Supplementary Information).

Diagnosis. The largest known haramiyid with a body mass estimated at 354 g; dental formula I1?-C0-P2-M2/i1-c0-p1-m2; the enlarged lower incisor fully covered with enamel; the only lower premolar high, triangular
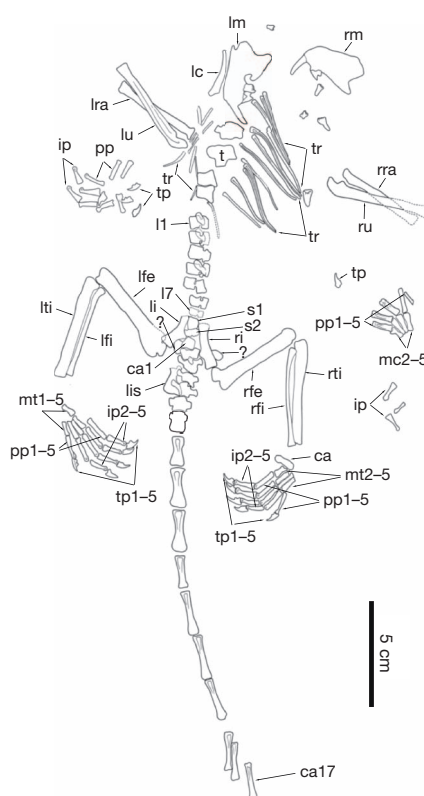
Figure 1 | The holotype specimen and line drawing of *Arboroharamiya jenkinsi* (STM33-9). The counterpart of the holotype is illustrated in Supplementary Fig. 1. ca, calcaneum; ca1–17, first to seventeenth caudal vertebrae; ip, intermediate phalanges; ip2–5, second to fifth intermediate phalanges; mc2–5, second to fifth metacarpals; mt1–5, second to fifth left metatarsals; mt2–5, second to fifth right metatarsals; l1–7, first to seventh lumbar vertebrae; lc, left clavicle; lfe, left femur; lfi, left fibula; li, left ilium; lis, left ischium; lm, left mandible; lra, left radius; lti, left tibia; lu, left ulna; pp, proximal phalanges; pp1–5, first to fifth proximal phalanges; rfe, right femur; rfi, right fibula; ri, right ilium; rm, right mandible; rra, right radius; rti, right tibia; ru, right ulna; t, thoracic vertebrae; tp, terminal phalanges; tp1–5, first to fifth terminal phalanges; tr, thoracic ribs; s1–2, first to second sacral vertebrae; ?, unknown element.

[1]Institute of Geology and Paleontology, Linyi University, Shuangling Road, Linyi City, Shandong 276005, China. [2]Shandong Tianyu Museum of Nature, Pingyi, Shandong 273300, China. [3]Key Laboratory of Vertebrate Evolution and Human Origin of Chinese Academy of Sciences, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing 100044, China. [4]Department of Biology, Indiana University of Pennsylvania, Indiana, Pennsylvania 15705, USA. [5]Division of Paleontology, American Museum of Natural History, Central Park West and 79th Street, New York, New York 10024, USA.

in lateral view and lacking serrations; the lower molar having a greatly inflated mesiolingual cusp (a1) and a central basin mesiodistally elongate and deeper distally than mesially; the upper dentition having at least a multi-cuspate incisor, probably two premolars and two molars; the distobuccal cusp (A1) the largest in the upper molars; the deep mandible lacking the postdentary trough and Meckelian groove but having a small angular process, a small coronoid process and a mandibular condyle that levels with the tooth row and is dorsoventrally oriented; limbs slim with proportionally short metapodials but long phalanges (Supplementary Information).

Arboroharamiya, as with other mammals, has body hair (preserved as impressions), a single-boned (dentary) mandible that implies a three-boned middle ear, and the digit formula of 2-3-3-3-3. The dentition is differentiated into incisors and multi-rooted premolars and molars, with the canine presumably lost. It differs from other mammals but is similar to allotherians in having two mesiodistally aligned rows of cusps that allow orthal (vertical) and palinal (backward), but not proal (forward) or transverse, jaw movement in mastication.

Arboroharamiya is also similar to other haramiyids in having basined molars with cusps of uneven heights, but differs from them in being larger and having a more inflated a1 and elongate central basin surrounded by more cusps in lower molars. *Arboroharamiya* further resembles multituberculates in having one pair of enlarged lower incisors, a multi-cuspate upper incisor, loss of the canine, two upper and lower molars, and the mandibular with a mesially extended masseteric fossa and a condyle positioned low and oriented more vertically than transversely. It differs from the Jurassic multituberculates in having a small angular process and a highly specialized dentition with one lower premolar and two upper premolars and with an occlusal pattern in which the enlarged a1 bites into the basin of the upper molar.

Previously, the only known haramiyid with the mandible and dentition preserved was *Haramiyavia*[10], which was thought to resemble *Morganucodon* and *Kuehneotherium* in having the masseteric fossa un-extended beyond the posterior part of the last molar, the condyle above the level of the teeth, and presence of the postdentary trough[2,5]. Although *Arboroharamiya* is similar to *Haramiyavia* in having a1 as the largest cusp and the lower molar basin deeper distally than mesially and possessing a gracile postcranial skeleton, it is morphologically more advanced than *Haramiyavia* in lacking the canine and having fewer incisors, premolars and molars. Most importantly, the dentary of *Arboroharamiya* is highly specialized and considerably different from that of *Haramiyavia*, but similar to those of derived multituberculates such as taeniolabidids, in being short and deep and in lacking the postdentary trough, which is one of the characteristics previously used to differentiate haramiyids from multituberculates[2,5,10].

Assignment of tooth locus and thus identification of species for most haramiyids are ambiguous because only isolated teeth are preserved[1–5]. The discovery of *Haramiyavia*[10] helped to solve part of the puzzle. However, because the upper premolars of *Haramiyavia* are not preserved, the difference of the upper premolar and molar of haramiyids remains unclear. In light of the new findings from *Arboroharamiya* (Fig. 2 and Supplementary Fig. 3), some haramiyid teeth identified as upper molars[4,11] are most likely premolars. Similarly, the largest cusp identified as b2 (mesiobuccal) in the lower molar of eleutherodontid haramiyids[2,11–13] is probably a1, and, if so, it would result in a different interpretation of the occlusal pattern in those haramiyids.

The occlusal pattern is a critical feature that was used not only to identify tooth locus and species within haramiyids[2,4,5], but more importantly to distinguish Allotheria from other Jurassic mammals that have transverse jaw movement during mastication[14]. In Allotheria, the upper and lower molariform teeth have essentially two longitudinal rows of cusps, and the buccal row of the lower molar was considered to bite into the valley between the two rows of the upper molar, involving orthal and/or palinal jaw movement[2–4]. The tooth morphology and wear pattern of *Arboroharamiya* confirm the orthal and palinal jaw movements in haramiyids, but nonetheless demonstrate that it is impossible
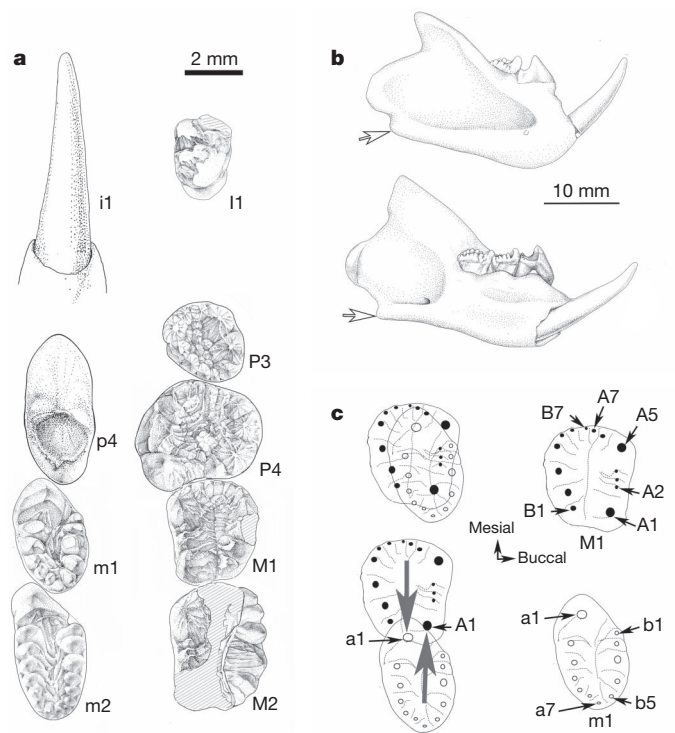

Figure 2 | Teeth, mandibles and tooth occlusal relationships of *Arboroharamiya jenkinsi*. a, Occlusal views of right upper and lower incisors (I/i), premolars (P/p) and molars (M/m). Some of the teeth (Supplementary Fig. 3) have been photographically flipped in the stippling drawings. The general shape of p4 is similar to that of *Kermackodon*, one of the earliest known multituberculates[4], but differs from it in lacking serrations. m1 and m2 are similar in having a high and inflated cusp a1 with cusps decreasing height distally. The distal end of the central basin is closed by cusps. Enamel ridges extend distally from cusps towards the basin, which enhance grinding as A1 of the upper molar 'moves' in the valley. The upper premolar differs from molars in being more rounded, with the broad central basin bearing numerous small cusps or crenulations. Upper molars are more mesiodistally elongate and have cusps A1 and A5 the largest and ridges extending mesially. Cusp B3 is the largest in cusp B-row. **b**, Buccal (top) and lingual (bottom) views of the mandible show the anterior extension of the masseteric fossa to the level below p4 and lack of the postdentary trough (Supplementary Fig. 2). The empty arrows point to the angular process. **c**, Line drawings illustrate the cusp numbering of M1 and m1 (right column) following ref. 2 and their relationship in occlusion (upper left). Grey arrows show the relative movements of a1 of m1 and A1 of M1 (bottom left). Tooth identification, measurements and photographs in Supplementary Information.

for the buccal row of the lower molar to bite into the valley of the upper molars. This is because the tall and inflated a1 at the longitudinal axis of m1 and the distally closed central basin in lower molars prevent such an occlusion. The cusp shape and arrangement, wear pattern and occlusal match of M1 and m1 show that, during mastication, a1 of the lower molar must have bit orthally in the basin of the upper molar in the puncture-crushing cycle and then moved palinally within the basin in a grinding cycle (Fig. 2 and Supplementary Figs 3–6). In a reversed symmetry, A1 of the upper molar bit into the central basin of the lower molar and 'moves' mesially in the valley of the lower molar. This 'double engaged' occlusion prevents both proal and transverse chewing motion; it creates wear in the tooth basin at the distal V-notch and on the buccal side of A1–3, but not the lingual side of M1. It also creates wear on the lingual and buccal sides of a1 in lower molars. This occlusal pattern is unique among mammals and differs from what has been interpreted for both haramiyids and multituberculates[2,4,10]. The tooth morphology and occlusal pattern suggest that *Arboroharamiya* is probably either granivorous, as proposed for some haramiyids[4], or omnivorous, as in multituberculates[3].

Postcranial features of *Arboroharamiya* show adaptation for an arboreal life. The femur head is not spherical but cylindrical, a derived feature present in omomyines and tarsiers[15]. *Arboroharamiya* has relatively short metapodials but long phalanges (Fig. 1 and Supplementary Fig. 7), unique among early Mesozoic mammals but characteristic of animals with prehensile hands and feet for arboreal life, such as arboreal didelphids and cheirogaleid primates[16] (Fig. 3). Even among arboreal mammals *Arboroharamiya* is distinctive in that both manual and pedal phalangeal indices, estimates for degree of prehensility of the hand-foot[17], are clearly higher than those of extant species[16,18] and extinct scansorial/arboreal species, such as *Eomaia* and *Sinodelphys* (Supplementary Figs 9 and 10). Moreover, the proximal caudal region is relatively long and bears expanded transverse processes. The transitional and longest caudal vertebrae are more distally positioned than those in nonprehensile taxa. These features are functionally related to the hypertrophy of the basal musculature necessary for increased gripping strength indicative of prehensile ability, as in some extant arboreal species[19,20]. The postcranial morphologies of *Arboroharamiya* suggest a gracile body for arboreal habitat preference.

The phylogenetic relationship of Allotheria remains controversial. It was presumed that haramiyids and multituberculates were not closely related[10], or they formed a clade in which multituberculates are derived from haramiyids, with the latter being paraphyletic[2,5,21]. Because of their unique dentition and early occurrences, allotherians were also considered to originate early from other mammals, even before mammaliaformes in the Triassic[2,4,5,22]. Our phylogenetic analysis supports the view that allotherians form a subgroup of the crown Mammalia[3,23,24] and that multituberculates are derived from haramiyids[2,5,21] (Fig. 4). Because *Haramiyavia*[10] and several other haramiyids[5,10] are from the Upper Triassic (the Norian–Rhaetic), the age for the origin of crown mammals would be in the time range of 228–201.3 Myr[25], younger than that estimated in ref. 7 but older than those in other studies[6,8]. With the ecological diversification recognized in early mammals[26], our phylogeny further implies that all major clades and feeding adaptations of mammals had diversified during the Jurassic, coincident with inferred diversifications of major lineages of insects[27,28] and angiosperms[29,30].

Owing to the fragmentary nature of most haramiyids, a thorough phylogenetic analysis of Allotheria remains impractical. However, *Arboroharamiya* demonstrates convincingly that haramiyids had become highly specialized in the Jurassic. It displays several mammalian features and fills some morphological gaps between *Haramiyavia* and multituberculates. Although morphological characteristics support allotherians as a clade, *Arboroharamiya* shows again that homoplasy is a common phenomenon within Mesozoic mammals[1]. Some features of *Arboroharamiya*, such as the reduced dentition—shared with advanced multituberculates—and elongated digits—shared with more advanced arboreal mammals—must be convergences. On the other hand, the dentition with multiple
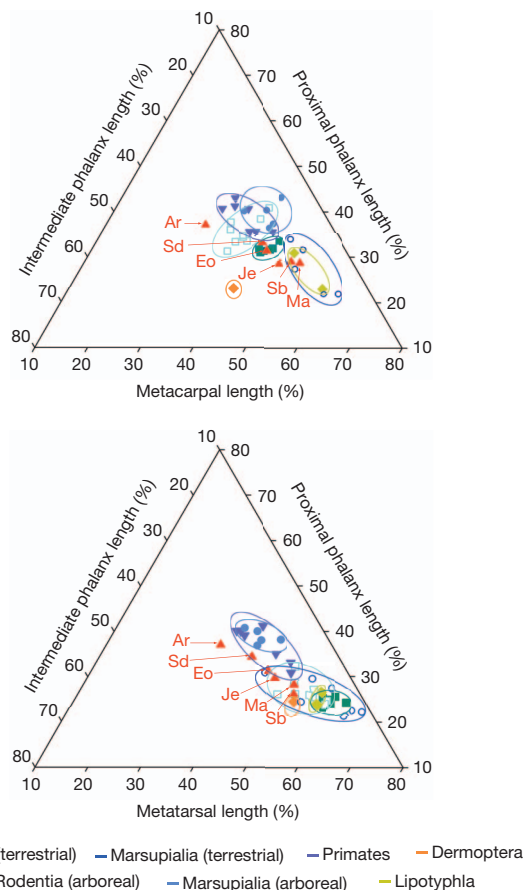
Figure 3 | Ternary diagrams showing intrinsic manual and pedal ray III proportions. Ternary plots showing relative metapodial, proximal and intermediate phalangeal lengths for the third digit ray of the hand and foot. The lengths of the third metapodial, proximal phalanx and intermediate phalanx are shown on their respective axes as a percentage of the combined length of the three segments. Compared to both fossil and extant taxa, *Arboroharamiya jenkinsi* has the intrinsic manual and pedal ray proportions typical of arboreal species in which the proximal and intermediate phalanges are long in respect to the metapodials (Fig. 1 and Supplementary Fig. 7). Abbreviations: Ar, *A. jenkinsi*; Eo, *Eomaia scansoria*; Je, *Jeholodens jenkinsi*; Ma, *Maotherium sinensis*; Sb, *Sinobaatar lingyuanensis*; Sd, *Sinodelphys szalayi*. Measurements and methods in part F of Supplementary Information.
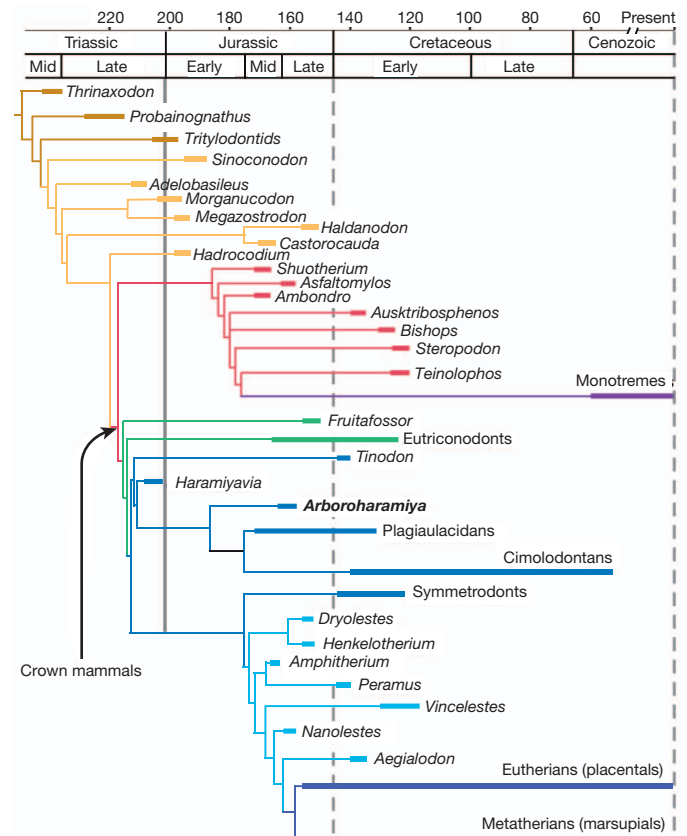

Figure 4 | Relationship of *Arboroharamiya* and geological distributions of major groups of Mesozoic mammals and their relatives. Thin lines represent the phylogenetic relationships and thick lines indicate geological distributions of the taxa. This is a simplified consensus tree (Supplementary Fig. 11) of 12 equally most parsimonious trees of PAUP (Phylogenetic Analysis Using Parsimony and Other Methods, version 4.0b), an analysis of 436 characters and 56 taxa with a focus on Mesozoic non-therian groups (modified from ref. 23; parts G–I in Supplementary Information). Test analyses for alternative hypotheses are in part J of Supplementary Information.

premolars in Jurassic multituberculates has to be considered as reversed from the condition of *Haramiyavia*. Regardless of various phylogenetic scenarios involving allotherians[2–5,10,21–24], morphological convergences and/or reversals were common in the early stage of mammalian evolution.

METHODS SUMMARY

Phylogenetic analyses were based on a data matrix consisting of 436 characters and 56 taxa (Supplementary Information), of which 389 characters are parsimony-informative, and were carried out with PAUP (version 4.0b). All characters were unordered and equally weighted, with gaps being treated as 'missing' and multi-state taxa interpreted as polymorphism. Character-states optimized as accelerated transformation (ACCTRAN).

Postcranial elements were measured using a digital caliper from 26 arboreal and non-arboreal extant mammals, except for *Caluromys*. Measurements for *Caluromys* and fossils, except for *Arboroharamiya*, were obtained from the literature. We estimated body mass using dimensions of the first lower molar, and conducted additional estimates based on the ulna, femur and tibia lengths, respectively. Full methods for body mass estimates, digit ray analyses and phylogenetic analyses are provided in Supplementary Information.

1. Sigogneau-Russell, D. Haramiyidae (Mammalia, Allotheria) en provenance du Trias supérieur de Lorraine (France). *Palaeontographica A* **206,** 137–198 (1989).
2. Butler, P. M. Review of the early allotherian mammals. *Acta Palaeontol. Pol.* **45,** 317–342 (2000).
3. Kielan-Jaworowska, Z., Cifelli, R. & Luo, Z.-X. *Mammals from the Age of Dinosaurs: Origins, Evolution, and Structure* (Columbia Univ. Press, 2004).
4. Butler, P. M. & Hooker, J. J. New teeth of allotherian mammals from the English Bathonian, including the earliest multituberculates. *Acta Palaeontol. Pol.* **50,** 185–207 (2005).
5. Hahn, G. & Hahn, R. Evolutionary tendencies and systematic arrangement in the Haramiyida (Mammalia). *Geol. Palaeontol.* **40,** 173–193 (2006).
6. Bininda-Emonds, O. R. et al. The delayed rise of present-day mammals. *Nature* **446,** 507–512 (2007).
7. Meredith, R. W. et al. Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science* **334,** 521–524 (2011).
8. O'Leary, M. A. et al. The placental mammal ancestor and the post-KPg radiation of placentals. *Science* **339,** 662–667 (2013).
9. Sullivan, C. et al. The Vertebrates of the Jurassic Daohugou Biota of Northeastern China. *J. Vertebr. Paleontol.* (in the press).
10. Jenkins, F. A., Gatesy, S. M., Shubin, N. H. & Amaral, W. W. Haramiyids and Triassic mammalian evolution. *Nature* **385,** 715–718 (1997).
11. Averianov, A. O., Lopatin, A. & Krasnolutskii, S. The First Haramiyid (Mammalia, Allotheria) from the Jurassic of Russia. *Dokl. Biol. Sci.* **437,** 103–106 (2011).
12. Kermack, K. A., Kermack, D. M., Lees, P. M. & Mills, J. R. New multituberculate-like teeth from the Middle Jurassic of England. *Acta Palaeontol. Pol.* **43,** 581–606 (1998).
13. Martin, T., Averianov, A. O. & Pfretzschner, H. U. Mammals from the Late Jurassic Qigu Formation in the southern Junggar Basin, Xinjiang, Northwest China. *Palaeobiodivers. Palaeoenviron.* **90,** 295–319 (2010).
14. Crompton, A. & Thomason, J. in *Functional Morphology in Vertebrate Paleontology* (ed. Thomason, J. J.) 55–75 (Cambridge Univ. Press, 1995).
15. Dagosto, M. & Schmid, P. Proximal femoral anatomy of omomyiform primates. *J. Hum. Evol.* **30,** 29–56 (1996).
16. Lemelin, P. Morphological correlates of substrate use in didelphid marsupials: implications for primate origins. *J. Zool.* **247,** 165–175 (1999).
17. Napier, J. *Hands* Ch. 2 26 (Princeton Univ. Press, 1993).
18. Kirk, E. C., Lemelin, P., Hamrick, M. W., Boyer, D. M. & Bloch, J. I. Intrinsic hand proportions of euarchontans and other mammals: implications for the locomotor behavior of plesiadapiforms. *J. Hum. Evol.* **55,** 278–299 (2008).
19. Schmitt, D., Rose, M. D., Turnquist, J. E. & Lemelin, P. Role of the prehensile tail during locomotion: experimental and osteological evidence. *Am. J. Phys. Anthropol.* **126,** 435–446 (2005).
20. Youlatos, D. Osteological correlates of tail prehensility in carnivorans. *J. Zool.* **259,** 423–430 (2003).
21. Hahn, G., Sigogneau-Russell, D. & Wouters, G. New data on Theroteinidae: their relations with Paulchoffatiidae and Haramiyidae. *Geol. Palaeontol.* **23,** 205–215 (1989).
22. Simpson, G. G. The principles of classification and a classification of mammals. *Bull. Am. Mus. Nat. Hist.* **85,** 1–350 (1945).
23. Luo, Z-X., Chen, P.-J., Li, G. & Chen, M. A new eutriconodont mammal and evolutionary development in early mammals. *Nature* **446,** 288–293 (2007).
24. Rowe, T., Rich, T. H., Vickers-Rich, P., Springer, M. & Woodburne, M. O. The oldest platypus and its bearing on divergence timing of the platypus and echidna clades. *Proc. Natl Acad. Sci. USA* **105,** 1238–1242 (2008).
25. Ogg, J. G. in *A Geological Time Scale 2012* (eds Gradstein, F. M., Ogg, J. G., Schmitz, M. D. & Ogg, G. M.) 681–730 (Elsevier, 2012).
26. Luo, Z.-X. Transformation and diversification in early mammal evolution. *Nature* **450,** 1011–1019 (2007).
27. Hunt, T. et al. A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. *Science* **318,** 1913–1916 (2007).
28. Wiegmann, B. M. et al. Episodic radiations in the fly tree of life. *Proc. Natl Acad. Sci. USA* **108,** 5690–5695 (2011).
29. Magallóan, S. A. & Sanderson, M. J. Angiosperm divergence times: the effect of genes, codon positions, and time constraints. *Evolution* **59,** 1653–1670 (2005).
30. Smith, S. A., Beaulieu, J. M. & Donoghue, M. J. An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proc. Natl Acad. Sci. USA* **107,** 5897–5902 (2010).

Supplementary Information is available in the online version of the paper.

Author Contributions X.Z., S.B. and J.M. designed the project. X.Z., S.B., X.W. and J.M. performed the research. S.B. and J.M. wrote the manuscript.

A latent capacity for evolutionary innovation through exaptation in metabolic systems

Aditya Barve[1,2] & Andreas Wagner[1,2,3]

Some evolutionary innovations may originate non-adaptively as exaptations, or pre-adaptations, which are by-products of other adaptive traits[1–5]. Examples include feathers, which originated before they were used in flight[2], and lens crystallins, which are light-refracting proteins that originated as enzymes[6]. The question of how often adaptive traits have non-adaptive origins has profound implications for evolutionary biology, but is difficult to address systematically. Here we consider this issue in metabolism, one of the most ancient biological systems that is central to all life. We analyse a metabolic trait of great adaptive importance: the ability of a metabolic reaction network to synthesize all biomass from a single source of carbon and energy. We use novel computational methods to sample randomly many metabolic networks that can sustain life on any given carbon source but contain an otherwise random set of known biochemical reactions. We show that when we require such networks to be viable on one particular carbon source, they are typically also viable on multiple other carbon sources that were not targets of selection. For example, viability on glucose may entail viability on up to 44 other sole carbon sources. Any one adaptation in these metabolic systems typically entails multiple potential exaptations. Metabolic systems thus contain a latent potential for evolutionary innovations with non-adaptive origins. Our observations suggest that many more metabolic traits may have non-adaptive origins than is appreciated at present. They also challenge our ability to distinguish adaptive from non-adaptive traits.

How evolutionary adaptations and innovations originate is one of the most profound questions in evolutionary biology. Previous work[1,2] emphasizes the importance of exaptations, also sometimes called pre-adaptations, for this origination. These are traits whose benefits to an organism are unrelated to the reasons for their origination; they are features that originally serve one (or no) function, and become later co-opted for a different purpose[1–5]. Although examples of exaptations occur from the macroscopic scale to the molecular[1–6] and abound also in human evolution[7], no number of examples could reveal how important exaptations are in the origination of adaptations in general. This limitation of case studies can be overcome in those biological systems where it is possible to study systematically many genotypes and the phenotypes they form[8–12].

One of these systems is metabolism. The metabolic genotype of an organism encodes a metabolic reaction network with hundreds of enzyme-catalysed chemical reactions. One of metabolism's fundamental tasks is to synthesize small biomass precursor molecules from environmental molecules, such as different organic carbon sources. An organism or metabolic network is said to be viable on a carbon source if it is able to synthesize all biomass molecules from this source. Viability on a new carbon source can be an important adaptation, and anecdotal evidence shows that this ability can originate as a pre-adaptation[13,14]. For example, laboratory evolution of *Pseudomonas putida* for increased biomass yield on xylose as a carbon source produces strains that utilize arabinose as efficiently as they do xylose, even though the ancestral strains did not utilize arabinose[14]. Thus, viability on arabinose can be a by-product of increased viability on xylose. We here analyse systematically whether such exaptations are typical or unusual in metabolic systems.

Our analysis relies on the ability to predict a metabolic phenotype from a metabolic genotype with the constraint-based method of flux balance analysis (Methods), to study not just one metabolic network but to explore systematically a vast space of possible metabolic networks. The members of this space can be described as follows. The currently known 'universe' of biochemical reactions comprises more than 5,000 chemical reactions with well-defined substrates and products. In the metabolic network of any one organism, however, only a fraction of these reactions take place, enabling us to describe this network through a binary presence/absence pattern of enzyme-catalysed reactions in the known reaction universe. Recent methods based on Markov chain Monte Carlo (MCMC) sampling (Methods) allow a systematic exploration of this space; that is, they permit the creation of arbitrarily large and uniform samples of networks with a given phenotype[12]. This sampling is based on long random walks through metabolic network space, where each step in a walk adds or eliminates a metabolic reaction from a metabolic network, with the only constraint being that the network remains viable on a focal carbon source. The starting point of the MCMC random walk is the *Escherichia coli* metabolic network, which we know a priori to be viable on different carbon sources[15]. Here we use this approach to create random samples of metabolic networks that are viable on a given set of carbon sources. We refer to such networks as random viable networks.

Our analysis focuses on 50 biologically relevant and common carbon sources[15] (Supplementary Table 1). For each carbon source C, we create a sample of 500 random viable networks that are viable on C if it is provided as the sole carbon source. We then use flux balance analysis to determine the viability of these networks on each of the 49 other carbon sources. This approach allows us to ask whether viability on C usually entails viability on other carbon sources. The answers to this and related questions show that potential exaptations are ubiquitous in metabolism.

We began our analysis with a sample of 500 random networks that were viable on glucose as the sole carbon source (Methods). Each network can synthesize the 63 essential biomass precursors of *E. coli*—many of which are important for most organisms[15,16]—in an aerobic minimal environment (Methods) containing glucose as the only carbon source. Importantly, we did not require that these 500 networks be viable on any carbon source except glucose.

We first examined whether these networks were viable on each of the 49 other carbon sources. The information resulting from this analysis can be represented, for each network, as a binary 'innovation vector' whose ith entry equals 1 if the network is viable on carbon source C_i, and equals 0 otherwise (Fig. 1a). We define the innovation index, $I_{Glucose}$, of a network to be the number of additional carbon sources on which each network is viable. The distribution of this index is shown in Fig. 1b. Ninety-six per cent of networks are viable on other carbon sources in addition to glucose ($I_{Glucose} > 0$). The mean innovation index

[1]Institute of Evolutionary Biology and Environmental Sciences, Building Y27, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland. [2]The Swiss Institute of Bioinformatics, Bioinformatics, Quartier Sorge, Bâtiment Genopode, 1015 Lausanne, Switzerland. [3]The Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA.
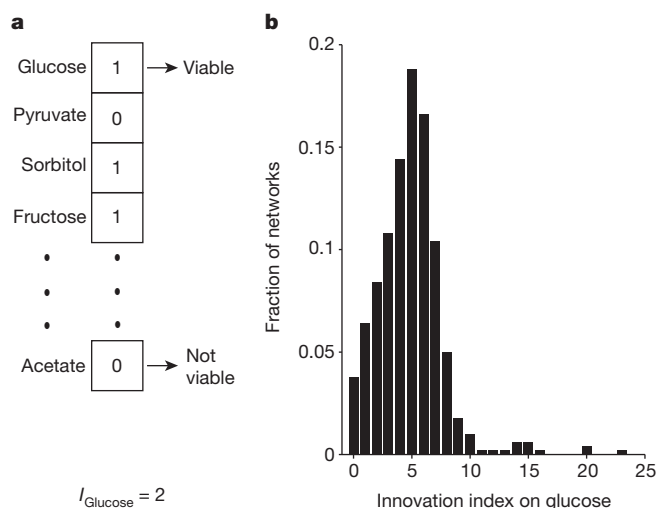
Figure 1 | Viability on glucose entails viability on multiple other carbon sources. a, The binary innovation vector of a hypothetical metabolic network that is viable on glucose. The vector shows that the random network is viable (labelled by 1) on glucose, sorbitol and fructose, but not viable (labelled by 0) on pyruvate and acetate. The innovation index of this network ($I_{Glucose} = 2$) is the number of additional carbon sources on which the network is viable. **b**, The distribution of innovation indices for 500 random networks viable on glucose. Only 4% of networks have $I_{Glucose} = 0$, meaning that they are viable only on glucose.

is $I_{Glucose} = 4.86$ (standard deviation, 2.83 carbon sources). This means that networks viable on glucose typically are also viable on almost 5 additional carbon sources. Ninety-four networks (18.8%) are viable on exactly 5 new carbon sources, and 187 networks (37.4%) are viable on 6 or more carbon sources. Viability on each such carbon source is a potential exaptation. This viability is merely a by-product of viability on glucose and could become an adaptation whenever this carbon source is the sole carbon source. We also found that different random viable networks differ in the additional carbon sources to which they are pre-adapted (Supplementary Figs 1 and 2). Most of the 50 carbon sources we study confer viability on at least one network in our sample (Supplementary Results). Moreover, a variation in our sampling procedures that allows only reactions already connected to a metabolism to be altered further increases the incidence of exaptation (Methods and Supplementary Fig. 3). Finally, complex metabolic networks that have more reactions have greater potential for exaptation (Supplementary Fig. 4).

We next asked whether the ability to grow on multiple additional carbon sources is a peculiarity of networks viable on glucose. To this end, we sampled, for each of our remaining 49 carbon sources, 500 random metabolic networks viable on this carbon source (for a total of $49 \times 500 = 24{,}500$ sampled networks). We then computed the distribution of the innovation index, I_C, for each carbon source C. Figure 2a shows the mean of this distribution (bars) and its coefficient of variation (vertical lines), that is, the ratio of the standard deviation to the mean. The figure shows that glucose (highlighted in red) is not unusual. Eighteen carbon sources (36%) have a greater average innovation index than glucose. For example, acetate allows viability on the greatest number (9.75) of additional carbon sources. Conversely, some carbon sources, such as adenosine ($I_{Adenosine} = 0.27$) and deoxyadenosine ($I_{Deoxyadenosine} = 0.1$), allow growth on fewer additional carbon sources than glucose. Carbon sources with a small average innovation index—entailing viability on few additional carbon sources—are also more variable in innovation index (Spearman's $\rho = -0.82, P < 10^{-101}$; see also Supplementary Fig. 5). Even though any one carbon source may confer growth on only few additional carbon sources in any one network (Fig. 2a), when considering all networks in a sample, it may still allow pre-adaptation to most other carbon sources (Supplementary Fig. 6).

In summary, viability on any one carbon source, C, usually entails viability on multiple other carbon sources, whose number and identity can vary with C. Viability on carbon sources never before encountered is thus a typical metabolic property. Environmental generalists capable of surviving on multiple carbon sources may be viable on many more carbon sources than occur in their environment (Supplementary Tables 2 and 3 and Supplementary Fig. 10).

We next asked whether metabolically close carbon sources show the highest potential for pre-adaptation. The centre path of Fig. 2b shows a hypothetical metabolic pathway that leads from one carbon source, C, to another, C_{new} (boxed area), and from there through (possibly multiple) further metabolic reactions to the synthesis of biomass. Figure 2c shows the same scenario, except that C and C_{new} are separated by several further reactions. It is possible that random networks viable on C are more likely to be viable also on C_{new} if C_{new} is closer to C, that is, if they are separated by fewer metabolic reactions, as in Fig. 2b. In this case, metabolite C_{new} may be less easy to bypass through an alternative pathway that originates somewhere between C and C_{new} (right-hand sequence of arrows in Fig. 2c).

To test this hypothesis (Supplementary Results), we analysed our 50 samples of 500 random metabolic networks, where networks in each sample were required to be viable on a different one of our 50 carbon sources. For each sample (carbon source C) and for each of the other 49 possible carbon sources, C_{new}, we asked whether the metabolic distance between C and C_{new} is correlated with the fraction of networks that are also viable on C_{new}. To answer this, we used metabolic networks that were selected for growth on C and were additionally viable on C_{new} (Methods). We then computed the mean metabolic distance and binned the distances. The results, pooled for all networks, are shown on the vertical axis of Fig. 2d, whose horizontal axis shows the mean metabolic distance (binned into nine bins). The closer C_{new} is to C, the more networks viable on C are also viable on C_{new} (Spearman's $\rho = -0.42, P = 10^{-87}, n = 1{,}990$). However, the figure also shows that the association is highly noisy, especially at low metabolic distances. Taking reaction irreversibility into account yields the same result (Spearman's $\rho = -0.39, P = 10^{-57}, n = 1{,}601$), as does a different way of computing distances between pairs of carbon sources (Methods and Supplementary Results). The association is noisy, because metabolism is highly reticulate (Supplementary Results).

Although metabolic 'nearness' cannot explain exaptations involving two carbon sources, biochemical similarities help explain why a network viable on C might be viable on one additional carbon source, C_{n1}, but not on another source, C_{n2}. Indeed, exaptations often involve carbon sources with broadly defined biochemical similarities (Supplementary Figs 7 and 8). For example, glycolytic carbon sources are more likely to entail exaptations for growth on other glycolytic carbon sources, and likewise for gluconeogenic carbon sources, as well as for carbon sources involved in nucleotide metabolism. Furthermore, we also find that pre-adaptation is synergistic; that is, the innovation index for a pair of carbon sources is greater than the sum of the innovation indices, I_{C1} and I_{C2} (Supplementary Fig. 9).

Our analysis has several limitations. First, it is based on present knowledge about the reaction universe. Future work may increase the number of known reactions, but this would not diminish, and could only enhance, the spectrum of possible exaptations. The reason is that additional reactions would allow the use of additional carbon sources by some metabolic networks. Second, most of our analysis focused on random networks that are viable on a specific carbon source, but selection in the wild can affect more than viability, which may affect the incidence of exaptations. Of special importance is selection that favours networks with a high rate of biomass synthesis. This particular selective constraint would not affect our conclusions, because we found that networks with high biomass synthesis rates have even greater potential for metabolic innovation than merely viable networks (Supplementary Table 4 and Supplementary Fig. 11). Third, we considered all necessary nutrient transporters to be present (Methods). If
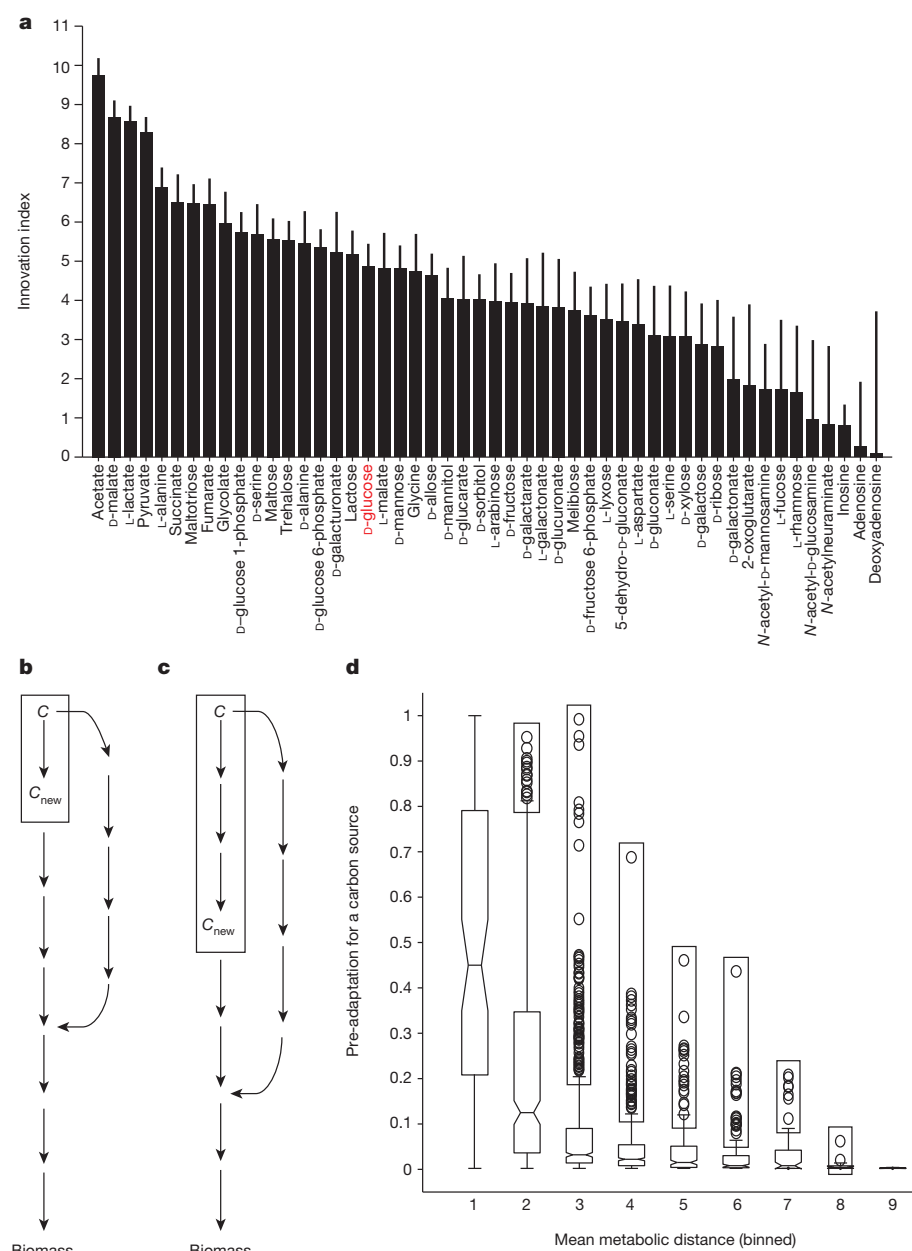
Figure 2 | Innovation varies with respect to the carbon source, C, and the mean metabolic distance between C and C_{new}. a, For each of 50 carbon sources (horizontal axis), the figure indicates the mean innovation index (bar) and its coefficient of variation (vertical line) for 500 random networks required to be viable on that carbon source. Note the broad distribution of the index. Some carbon sources, such as acetate, allow viability on more than nine additional carbon sources, on average, whereas others, such as deoxyadenosine, support viability on fewer than one additional carbon source. The innovation index of glucose (red) is typical compared with other carbon sources. **b**, A hypothetical carbon source, C_{new}, which can be synthesized from another carbon source, C, in one reaction (arrow), and which leads, through multiple further reactions, to the synthesis of biomass. Some metabolic networks may have an alternative metabolic pathway that bypasses C_{new} altogether (right-hand sequence of arrows). **c**, Like **b**, but with C_{new} and C separated by multiple reactions. The fewer reactions separate C and C_{new}, the more likely it is that C_{new} is not bypassed by some alternative metabolic pathway, and that viability on C therefore implies viability on C_{new}. **d**, Testing the hypothesis in **c**. The horizontal axis shows the mean number of reactions that separate C and C_{new} in networks that are viable on both C and C_{new}, binned into integer intervals according to the floor of this number (that is, the greatest smaller integer). The vertical axis shows the fraction of random metabolic networks required to be viable on carbon source C that are additionally viable on C_{new}. We note that the potential for innovation decreases with increasing distance. Box edges, 25th and 75th percentiles; central horizontal line in each box, median; whiskers, ± 2.7 s.d.; open circles, outliers. Data are based on samples of 500 random viable networks for each of 50 carbon sources C ($n = 25,000$).

this is not the case, the incidence of exaptation may be reduced. In this regard, we note that 84% of *E. coli* transporters can transport multiple molecules[17] and that their substrate specificity can change rapidly[18], thus ameliorating this constraint. Fourth, real metabolic networks may contain more reactions connected to the rest of metabolism than do our randomly sampled networks. However, when restricting our analysis to networks in which all reactions are connected, we found an even greater incidence of exaptation than in random networks (Methods, Supplementary Results and Supplementary Fig. 3). Thus, our results provide a lower bound on the incidence of exaptations. Finally, most of our analysis is based on sampling a limited number of 500 networks viable on each carbon source, but sampling 5,000 random networks for select carbon sources yielded identical results (Supplementary Fig. 12).

Our observations show that latent metabolic abilities are pervasive features of carbon metabolism. They expose non-adaptive origins of potentially useful carbon-source utilization traits as a universal and inevitable feature of metabolism. The abundance of non-adaptive trait origins results from the complexity of metabolic systems, which have many enzyme parts that can jointly form multiple metabolic phenotypes,

but this ability is not restricted to metabolic networks. Many enzymes are capable of using various substrates[17,19], which can further increase network complexity and the potential for exaptation. The ability to form multiple phenotypes also occurs in regulatory circuits[20], which can form different patterns of molecular activity, as well as in RNA molecules[21], which can form multiple conformations with different biological functions. Systematic analyses of genotype–phenotype relationships are becoming increasingly possible in such systems[22,23], and already hint at exaptive origins of molecular traits. If confirmed in systematic analyses like ours, the pervasiveness of non-adaptive traits may require a rethinking of the early origins of beneficial traits.

METHODS SUMMARY

We used MCMC random walks that utilize reaction swapping to sample random viable metabolic networks[12], and used flux balance analysis[24] to compute the viability of metabolic networks during the MCMC procedure. We performed all analyses for minimal aerobic growth environments composed of a sole carbon source, along with oxygen, ammonium, inorganic phosphate, sulphate, sodium, potassium, cobalt, iron (Fe^{2+} and Fe^{3+}), protons, water, molybdate, copper, calcium, chloride, magnesium, manganese and zinc[15].

1. Darwin, C. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* 6th edn (Murray, 1872).
2. Gould, S. J. & Vrba, E. S. Exaptation: a missing term in the science of form. *Paleobiology* **8**, 4–15 (1982).
3. True, J. R. & Carroll, S. B. Gene co-option in physiological and morphological evolution. *Annu. Rev. Cell Dev. Biol.* **18**, 53–80 (2002).
4. Zákány, J. & Duboule, D. Hox genes in digit development and evolution. *Cell Tissue Res.* **296**, 19–25 (1999).
5. Keys, D. N. *et al.* Recruitment of a hedgehog regulatory circuit in butterfly eyespot evolution. *Science* **283**, 532–534 (1999).
6. Tomarev, S. I. & Piatigorsky, J. Lens crystallins of invertebrates. Diversity and recruitment from detoxification enzymes and novel proteins. *Eur. J. Biochem.* **235**, 449–465 (1996).
7. Pievani, T. & Serrelli, E. Exaptation in human evolution: how to test adaptive vs exaptive evolutionary hypotheses. *J. Anthropol. Sci.* **89**, 9–23 (2011).
8. Schuster, P., Fontana, W., Stadler, P. F. & Hofacker, I. L. From sequences to shapes and back: a case study in RNA secondary structures. *Proc. R. Soc. Lond. B* **255**, 279–284 (1994).
9. Lipman, D. J. & Wilbur, W. J. Modelling neutral and selective evolution of protein folding. *Proc. R. Soc. Lond. B* **245**, 7–11 (1991).
10. Cowperthwaite, M. C., Economo, E. P., Harcombe, W. R., Miller, E. L. & Meyers, L. A. The ascent of the abundant: how mutational networks constrain evolution. *PLoS Comput. Biol.* **4**, e1000110 (2008).
11. Ferrada, E. & Wagner, A. A comparison of genotype-phenotype maps for RNA and proteins. *Biophys. J.* **102**, 1916–1925 (2012).
12. Samal, A., Matias Rodrigues, J. F., Jost, J., Martin, O. C. & Wagner, A. Genotype networks in metabolic reaction spaces. *BMC Syst. Biol.* **4**, 30 (2010).
13. Poulsen, T. S., Chang, Y.-Y. & Hove-Jensen, B. D-allose catabolism of *Escherichia coli*: involvement of alsl and regulation of als regulon expression by allose and ribose. *J. Bacteriol.* **181**, 7126–7130 (1999).
14. Meijnen, J.-P., De Winde, J. H. & Ruijssenaars, H. J. Engineering Pseudomonas putida S12 for efficient utilization of D-xylose and L-arabinose. *Appl. Environ. Microbiol.* **74**, 5031–5037 (2008).
15. Feist, A. M. *et al.* A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.* **3**, 121 (2007).
16. Neidhardt, F. & Ingraham, J. *Escherichia Coli and Salmonella Typhimurium: Cellular and Molecular Biology* Vol. 1, 13–16 (American Society for Microbiology, 1987).
17. Nam, H. *et al.* Network context and selection in the evolution to enzyme specificity. *Science* **337**, 1101–1104 (2012).
18. Aguilar, C. *et al.* Genetic changes during a laboratory adaptive evolution process that allowed fast growth in glucose to an *Escherichia coli* strain lacking the major glucose transport system. *BMC Genomics* **13**, 385 (2012).
19. Kim, J., Kershner, J. P., Novikov, Y., Shoemaker, R. K. & Copley, S. D. Three serendipitous pathways in *E. coli* can bypass a block in pyridoxal-5′-phosphate synthesis. *Mol. Syst. Biol.* **6**, 436 (2010).
20. Martin, O. C. & Wagner, A. Multifunctionality and robustness trade-offs in model genetic circuits. *Biophys. J.* **94**, 2927–2937 (2008).
21. Ancel, L. W. & Fontana, W. Plasticity, evolvability, and modularity in RNA. *J. Exp. Zool.* **288**, 242–283 (2000).
22. Amitai, G., Gupta, R. D. & Tawfik, D. S. Latent evolutionary potentials under the neutral mutational drift of an enzyme. *HFSP J.* **1**, 67–78 (2007).
23. Isalan, M. *et al.* Evolvability and hierarchy in rewired bacterial gene networks. *Nature* **452**, 840–845 (2008).
24. Price, N. D., Reed, J. L., Palsson, B. & Ø. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Rev. Microbiol.* **2**, 886–897 (2004).

METHODS

Flux balance analysis. Flux balance analysis (FBA) is a constraint-based computational method[24,25] used to predict synthetic abilities and other properties of large metabolic networks, which are complex systems of enzyme-catalysed chemical reactions. FBA requires information about the stoichiometry of each molecular species participating in the chemical reactions of a metabolic network. This stoichiometric information is represented as a stoichiometric matrix, S, of dimensions $m \times n$, where m denotes the number of metabolites and n denotes the number of reactions in a network[24,25]. FBA also assumes that the network is in a metabolic steady state, such as would be attained by an exponentially growing microbial population in an unchanging environment. This assumption makes it possible to impose the constraint of mass conservation on the metabolites in the network. This constraint can be expressed as $Sv = 0$, where v denotes a vector of metabolic fluxes whose entries, v_i, describe the rate at which reaction i proceeds. The solutions, or 'allowable' fluxes, of this equation form a large solution space, but not all of these solutions may be of biological interest. To restrict this space to fluxes of interest, FBA uses linear programming to maximize a biologically relevant quantity in the form of a linear objective function Z (ref. 25). Specifically, the linear programming formulation of an FBA problem can be expressed as

$$\max\{Z\} = \max\{c^{\mathrm{T}}v \mid Sv = 0, a \le v \le b\}$$

The vector c contains a set of scalar coefficients that represent the maximization criterion, and the individual entries of vectors a and b respectively contain the minimal and maximal possible fluxes for each reaction in v; that is, each entry v_i is bounded from below by a_i and bounded from above by b_i.

We are here interested in predicting whether a metabolic network can sustain life in a given spectrum of environments, that is, whether it can synthesize all necessary small biomass molecules (biomass precursors) required for survival and growth. In a free-living bacterium such as *E. coli*, there are more than 60 such molecules, which include 20 proteinaceous amino acids, DNA and RNA nucleotide precursors, lipids and cofactors. We use the *E. coli* biomass composition[15] to define the objective function and the vector c, because most molecules in *E. coli*'s biomass would be typically found in free-living organisms. We used the package CLP (1.4, Coin-OR; https://projects.coin-or.org/Clp) to solve the linear programming problems mentioned above.

Chemical environments. Along with the biomass composition and stoichiometric information about a metabolic network, it is necessary to define one or more chemical environments that contain the nutrients needed to synthesize biomass precursors. Here we consider only minimal aerobic growth environments composed of a sole carbon source, along with oxygen, ammonium, inorganic phosphate, sulphate, sodium, potassium, cobalt, iron (Fe^{2+} and Fe^{3+}), protons, water, molybdate, copper, calcium, chloride, magnesium, manganese and zinc[15]. When studying the viability of a metabolic network in different environments, we vary the carbon source while keeping all other nutrients constant. When we say, for example, that a particular network is viable on 20 carbon sources, we mean that the network can synthesize all biomass precursors when each of these carbon sources is provided as the sole carbon source in a minimal medium. For reasons of computational feasibility, we restrict ourselves to 50 carbon sources (Supplementary Table 1). They are all carbon sources on which *E. coli* is known to be viable from experiments[15]. We chose these carbon sources because many of them are prominent, and because they are of known biological relevance, but we emphasize that our observations do not otherwise make a statement about the metabolism of *E. coli* or its close relatives. They apply to metabolic networks that vary much more broadly in reaction composition than any relative of *E. coli*, because of our network sampling approach described below, which effectively randomizes the reaction composition of a microbial metabolism.

The known reaction universe. The known reaction universe is a list of metabolic reactions known to occur in some organisms. For the construction of this universe, we used data from the LIGAND database[26,27] of the Kyoto Encyclopedia of Genes and Genomes[28,29]. The LIGAND database is divided into two subsets—the REACTION database and the COMPOUND database. These two databases together provide information about metabolic reactions, participating chemical compounds and associated stoichiometric information in an interlinked manner.

As we described earlier[12,30,31], we specifically used the REACTION and COMPOUND databases to construct our universe of reactions while excluding all reactions involving polymer metabolites of unspecified numbers of monomers, or general polymerization reactions with uncertain stoichiometry; reactions involving glycans, owing to their complex structure; reactions with unbalanced stoichiometry; and reactions involving complex metabolites without chemical information[29]. The published *E. coli* metabolic model (*iAF1260*) consists of 1,397 non-transport reactions[15]. We merged all reactions in the *E. coli* model with

the reactions in the LIGAND database, and retained only the non-duplicate reactions. After these procedures of pruning and merging, our universe of reactions consisted of 5,906 non-transport reactions and 5,030 metabolites.

Sampling of random viable metabolic networks. In an organism, a metabolic network can change through mutations. They can lead to addition of new reactions, by way of horizontal gene transfer, or through the evolution of enzymes with novel activities. They can also lead to loss of reactions through loss-of-function mutations in enzyme-coding genes. Natural selection can preserve those changed metabolic networks that are viable in a particular environment. Together, mutational processes and selection may change a metabolic network drastically on a long evolutionary timescale. Recent work has shown that even metabolic networks that differ greatly in their sets of reactions can have the same metabolic phenotype, that is, the same biosynthetic ability[32]. We here use a recently developed MCMC random-sampling[12,30,31,33,34] procedure to generate metabolic networks that are viable in specific environments, but that contain an otherwise random complement of metabolic reactions. Briefly, this procedure involves random walks in the space of all possible networks. During any one such random walk, a metabolic network can change through the addition and deletion of reactions. Although this process resembles the biological evolution of metabolic networks through horizontal gene transfer and (recombination-driven) gene deletions, we here use it for the sole purpose of creating random samples of metabolic networks from the space of all such networks[12,34].

In any one MCMC random walk, we keep the total number of reactions at the same number as in the starting *E. coli* network (1,397; ref. 15), to avoid artefacts due to varying reaction network size[12]. Specifically, each mutation step in a random walk involves the addition of a randomly chosen reaction from the reaction universe, followed by the deletion of a randomly chosen metabolic reaction from the metabolic network. We call such a sequence of reaction addition and deletion a reaction swap. Reaction addition does not abolish the viability of a network in any environment. However, reaction deletion might. Thus, after a reaction deletion, we use FBA to ask whether the network is still viable, that is, whether it can synthesize all biomass precursors, in the specified environment. If so, we accept the deletion; otherwise, we reject it and choose another reaction for deletion at random, until we have found a deletion that retains viability. After that, we accept the reaction swap, thus completing a single step in the random walk. We do not subject transport reactions to reaction swaps. These reactions are therefore present in all networks generated by our random walk.

Any MCMC random walk begins from a single starting network, in our case that of *E. coli*. The theory behind MCMC sampling[12,34] shows that it is important to carry out as many reaction swaps as possible for MCMC to 'erase' the random walker's similarity ('memory') to the initial network. The reason is that successive genotypes in a random walk are strongly correlated in their properties, because they differ by only one reaction pair. These correlations decrease as the number of reaction swaps increases. Because we are interested in analysing growth phenotypes of networks, correlations to the initial network would result in identification of growth on carbon sources similar to those of the starting network. In past work[12,30], we found that for the network sizes that we use (1,397 reactions), 3×10^3 reactions swaps are sufficient to erase the similarity of the final network to the starting network. To err on the side of caution, we thus carried out 5×10^3 reaction swaps before beginning to sample, and sampled a network every 5×10^3 reaction swaps thereafter. In this way, we generated samples of 500 random viable metabolic networks through an MCMC random walk of 2.5×10^6 reaction swaps. We carried out different random walks to sample networks viable on different carbon sources.

For some of our analyses, we also sampled random metabolic networks of sizes different from that of the *E. coli* metabolic network. To do this, we followed a previously established procedure[12,30,31] to create a starting network for an MCMC random walk that has the desired size. This procedure first converts the known universe of reactions into a 'global' metabolic network by including the *E. coli* transport reactions in it. Not surprisingly, this global network can produce all biomass components and is therefore viable on all carbon sources studied here. We used this global network to delete successively a sequence of randomly chosen reactions in the following way. After each reaction deletion, FBA was used to determine whether the network was still viable on a given carbon source. If so, the deletion was accepted; otherwise, another reaction was chosen at random for deletion. We deleted in this way as many reactions as needed to generate a network of the desired size. We then used this network as the starting network for an MCMC random walk, as described above, to generate samples of 500 random viable networks.

Identification of disconnected non-functional reactions. We performed some of our analysis with a version of the reaction universe that does not contain disconnected reactions. Reactions that are not connected to the rest of a metabolic network would be non-functional, because they cannot carry a non-zero steady-state

metabolic flux, and thus could not contribute to the synthesis of biomass. The genes encoding them would eventually be lost from a genome. (We note that this loss could still take tens of thousands of years, given known deleterious mutation rates and generation times[35,36], which is enough for some for other genetic or environmental changes to render these reactions functional.) We define a disconnected reaction as a reaction that does not share any one substrate or any one product with any other reaction in the known reaction universe. We focus here on reactions in the universe rather than in one metabolic network, because an individual network can gain additional reactions that may connect previously disconnected reactions. We note that even this 'universal' definition of disconnectedness depends on our current knowledge of biochemistry, as well as on the environment, because the right environment could supply metabolites that connect previously disconnected reactions or pathways to the rest of a metabolic network. To identify the connected universe, we removed disconnected reactions. Because this removal may render other reactions disconnected, we repeated this process iteratively until no further reactions in the universe became disconnected. In this way, we found that 3,646 of the 5,906 reactions in the universe of reactions were connected. We used this connected universe in some analyses to generate network samples using the MCMC approach.

Estimation of the metabolic distance between carbon sources. To compute the metabolic distance between a pair of carbon sources, C and C_{new}, we used the 500 networks selected for growth on a specific carbon source, C. We first represented a network as a substrate graph[37]. In this graph, vertices correspond to metabolites. Two metabolites (vertices) are linked by an edge if the metabolites participate in the same metabolic reaction, be it as an educt or as a product. We excluded 'currency' metabolites from this substrate graph, which are metabolites that transfer small chemical groups and are involved in many reactions[38]. Specifically, we excluded protons, H_2O, ATP (adenosine triphosphate), ADP (adenosine diphosphate), AMP (adenosine monophosphate), NADP(H) (nicotinamide adenine dinucleotide diphosphate), NAD(H) (nicotinamide adenine dinucleotide), and P_i (inorganic phosphate), CoA (coenzyme A), hydrogen peroxide, ammonia, ammonium, bicarbonate, GTP (guanosine triphosphate), GDP (guanosine diphosphate), and PP_i (inorganic diphosphate) that occurred in both the cytoplasmic and periplasmic compartments[15]. In addition, we excluded oxidized and reduced forms of cofactors such as quinone, ubiquinone, glutathione, thioredoxin, flavodoxin and flavin mononucleotide. That is, we eliminated all vertices corresponding to these metabolites when constructing the substrate graph. For each metabolic network, we constructed two substrate graphs: one in which the reaction irreversibility was ignored and all reactions were considered reversible, and one in which irreversibility was taken into account. For a network selected for growth on carbon source C, we calculated the shortest distance from C to each exapted carbon source, C_{new}, in the substrate graph of that network, as computed by a breadth-first search[39]. We preformed this analysis for each network in our ensemble of 500 networks viable on C. The distance between C and C_{new} was then computed as the mean of the metabolic distances based on networks viable on both carbon sources.

We also computed the metabolic distance for any two carbon sources by representing the universe of reactions as a graph in the above manner. We again constructed two substrate graphs, as above. Taking irreversibility into account increases the maximal distance to infinity because some carbon sources are connected by irreversible reactions.

Clustering of carbon sources based on the innovation matrix. Entry I_{ij} of the innovation matrix, I, represents the fraction of random metabolic networks that we required to be viable on carbon source C_i and that was additionally viable on carbon source C_j. To cluster the entries of this matrix, we first computed for all pairs of rows in this matrix the quantity $d = 1 - \rho$, where ρ is the Spearman rank correlation coefficient between the row entries. This yielded a new distance matrix that describes the distances between all pairs of rows. We clustered the rows of I by applying UPGMA (unweighted pair group method with arithmetic means[40]), a hierarchical clustering method, to the distance matrix.

Hierarchical clustering with UPGMA classifies data such that the average distance between elements belonging to the same cluster is lower than the average distance between elements belonging to different clusters[12]. UPGMA identified two clusters of glycolytic and gluconeogenic carbon sources, and we wanted to know whether the distances between them were significantly different. To this end, we first calculated the distribution of distances $d = 1 - \rho$ for all pairs of row vectors of I within each of the two clusters. We called the resulting distance distribution the 'within-cluster' distance distribution. Similarly, we computed the distances between any pair of row vectors belonging to two different clusters. These formed a 'between-cluster' distance distribution. We then used the non-parametric Mann–Whitney U-test to check whether these two distributions were significantly different.

Estimation of carbon waste production. FBA determines the maximal biomass yield achievable by a network for a given carbon source[25]. However, even when a network produces the maximally achievable yield, not all of the carbon input into the network may be converted into biomass. The non-converted carbon input constitutes carbon waste. Such unused carbon can be secreted in the form of one or more metabolites. For example, in a glucose minimal environment *E. coli* secretes carbon dioxide and acetate into the extracellular compartment as carbon waste. FBA estimates the amount of each metabolite secreted per unit time[15,25]. To estimate the amount of carbon waste that a random network viable on glucose produces, we first identified the different metabolites that it secretes as waste and then computed the amount of carbon waste per metabolite as the product of carbon atoms in that metabolite and the amount of the metabolite secreted (millimoles per gram dry weight per hour). The total carbon waste produced by a network is computed as the sum of the above quantity for each of the secreted carbon-containing molecules. We repeated the above procedure for each random network in a sample of 500 random networks viable on glucose. We found a total of 62 metabolites that are secreted as waste metabolites in at least one network of our sample of networks viable on glucose.

We carried out all numerical analyses using MATLAB (Mathworks Inc.).

25. Kauffman, K. J., Prakash, P. & Edwards, J. S. Advances in flux balance analysis. *Curr. Opin. Biotechnol.* **14,** 491–496 (2003).
26. Goto, S., Nishioka, T. & Kanehisa, M. LIGAND: chemical database of enzyme reactions. *Nucleic Acids Res.* **28,** 380–382 (2000).
27. Goto, S., Okuno, Y., Hattori, M., Nishioka, T. & Kanehisa, M. LIGAND: database of chemical compounds and reactions in biological pathways. *Nucleic Acids Res.* **30,** 402–404 (2002).
28. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28,** 27–30 (2000).
29. Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M. & Hirakawa, M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **38,** D355–D360 (2010).
30. Barve, A., Rodrigues, J. F. M. & Wagner, A. Superessential reactions in metabolic networks. *Proc. Natl Acad. Sci. USA* **109,** E1121–E1130 (2012).
31. Matias Rodrigues, J. F. & Wagner, A. Evolutionary plasticity and innovations in complex metabolic reaction networks. *PLoS Comput. Biol.* **5,** e1000613 (2009).
32. Henry, C. S. *et al.* High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nature Biotechnol.* **28,** 977–982 (2010).
33. Matias Rodrigues, J. F. & Wagner, A. Genotype networks, innovation, and robustness in sulfur metabolism. *BMC Syst. Biol.* **5,** 39 (2011).
34. Binder, K. & Heerman, D. W. *Monte Carlo Simulation in Statistical Physics* (Springer, 2010).
35. Koskiniemi, S., Sun, S., Berg, O. G. & Andersson, D. I. Selection-driven gene loss in bacteria. *PLoS Genet.* **8,** e1002787 (2012).
36. Ochman, H., Elwyn, S. & Moran, N. A. Calibrating bacterial evolution. *Proc. Natl Acad. Sci. USA* **96,** 12638–12643 (1999).
37. Wagner, A. & Fell, D. A. The small world inside large metabolic networks. *Proc. R. Soc. Lond. B* **268,** 1803–1810 (2001).
38. Ma, H.-W. & Zeng, A.-P. The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics* **19,** 1423–1430 (2003).
39. Moore, E. in *Proc. Internat. Symp. Theory Switching, Ann. Comput. Lab. Harvard Univ.* 285–292 (Harvard Univ. Press, 1959).
40. Sokal, R. R. & Michener, C. D. A statistical method for evaluating systematic relationships. *Univ. Kansas Sci. Bull.* **28,** 1409–1438 (1958).

LETTER

The haplotype–resolved genome and epigenome of the aneuploid HeLa cancer cell line

Andrew Adey[1]*, Joshua N. Burton[1]*, Jacob O. Kitzman[1]*, Joseph B. Hiatt[1], Alexandra P. Lewis[1], Beth K. Martin[1], Ruolan Qiu[1], Choli Lee[1] & Jay Shendure[1]

The HeLa cell line was established in 1951 from cervical cancer cells taken from a patient, Henrietta Lacks. This was the first successful attempt to immortalize human-derived cells *in vitro*[1]. The robust growth and unrestricted distribution of HeLa cells resulted in its broad adoption—both intentionally and through widespread cross-contamination[2]—and for the past 60 years it has served a role analogous to that of a model organism[3]. The cumulative impact of the HeLa cell line on research is demonstrated by its occurrence in more than 74,000 PubMed abstracts (approximately 0.3%). The genomic architecture of HeLa remains largely unexplored beyond its karyotype[4], partly because like many cancers, its extensive aneuploidy renders such analyses challenging. We carried out haplotype-resolved whole-genome sequencing[5] of the HeLa CCL-2 strain, examined point- and indel-mutation variations, mapped copy-number variations and loss of heterozygosity regions, and phased variants across full chromosome arms. We also investigated variation and copy-number profiles for HeLa S3 and eight additional strains. We find that HeLa is relatively stable in terms of point variation, with few new mutations accumulating after early passaging. Haplotype resolution facilitated reconstruction of an amplified, highly rearranged region of chromosome 8q24.21 at which integration of the human papilloma virus type 18 (HPV-18) genome occurred and that is likely to be the event that initiated tumorigenesis. We combined these maps with RNA-seq[6] and ENCODE Project[7] data sets to phase the HeLa epigenome. This revealed strong, haplotype-specific activation of the proto-oncogene *MYC* by the integrated HPV-18 genome approximately 500 kilobases upstream, and enabled global analyses of the relationship between gene dosage and expression. These data provide an extensively phased, high-quality reference genome for past and future experiments relying on HeLa, and demonstrate the value of haplotype resolution for characterizing cancer genomes and epigenomes.

We generated a haplotype-resolved genome sequence of HeLa CCL-2 using a multifaceted approach that included shotgun, mate-pair and long-read sequencing, as well as sequencing of pools of fosmid clones[5] (Supplementary Table 1). To catalogue variants, we carried out conventional shotgun sequencing to 88× non-duplicate coverage and reanalysed 11 control germline genomes in parallel[8] (Supplementary Tables 2 and 3). Although normal tissue corresponding to HeLa is unavailable, the total number of single-nucleotide variants (SNVs) identified in HeLa CCL-2 ($n = 4.1 \times 10^6$) and the proportion overlapping with the 1000 Genomes Project[9] (90.2%) were similar to controls (mean $n = 4.2 \times 10^6$ and 87.7%, respectively), suggesting that HeLa has not accumulated appreciably large numbers of somatic SNVs relative to inherited variants. Indel variation was unremarkable after accounting for differences in coverage (Supplementary Fig. 1). Short tandem repeat profiles of HeLa also resembled controls, consistent with mismatch repair proficiency (Supplementary Fig. 2).

After removing protein-altering variants that overlapped with the 1000 Genomes Project or the Exome Sequencing Project[10], similar numbers of private protein-altering (PPA) SNVs were found in HeLa ($n = 269$) and controls (mean $n = 391$). Gene ontology analysis found that all terms enriched for PPA variants in HeLa ($P \leq 0.01$) were also enriched in at least one control (except for 'startle response' in HeLa), suggesting that known cancer-related pathways are not perturbed extensively by point or indel mutations (Supplementary Fig. 3). Although a previous study of the HeLa transcriptome[11] reported an enrichment of putative mutations in cell-cycle- and E2F-related genes, subsequently generated population-scale data sets contain all variants that we observed in these genes, suggesting that they are inherited and benign rather than somatic and pathogenic.

The overlap between PPA variants and the Catalogue of Somatic Mutations in Cancer (COSMIC)[12] was similar for HeLa ($n = 1$) and control genomes (mean $n = 2.6$). The gene-level overlap with the Sanger Cancer Gene Census (SCGC)[12] was also similar for HeLa ($n = 4$) and control genomes (mean $n = 8.7$). Canonical tumour suppressors and oncogenes were notably absent among the five SCGC genes with PPA variants in HeLa (*BCL11B* (B-cell CLL/lymphoma 11B (zinc finger protein)), *EP300* (E1A binding protein p300), *FGFR3* (fibroblast growth factor receptor 3), *NOTCH1* and *PRDM16* (PR domain containing 16), Supplementary Tables 3–6). However, three are associated with HPV-mediated oncogenesis (*FGFR3, EP300, NOTCH1*) and may be ancillary to the dominant role of HPV oncoproteins in HeLa and other HPV⁺ cervical carcinomas[13]. Mutations in *FGFR3* have been noted previously in cervical carcinomas, although infrequently and at different residues than observed here[14]. Both *EP300* and *NOTCH1* are recurrently mutated in diverse cancers and are involved in Notch signalling, a pathway that is dysregulated in HeLa[15]. *EP300*, which encodes the transcriptional co-activator p300, interacts directly with viral oncoproteins such as HPV-16 E6 and HPV-16 E7 (ref. 16). Although the in-frame deletion of a highly conserved amino acid in *EP300* seems to be somatic (heterozygous within a loss-of-heterozygosity (LOH) region), it is still possible that the others are rare, inherited variants or passenger mutations. Further studies are required to resolve their functional relevance and to assess whether these genes are recurrently altered in HPV⁺ cervical carcinomas.

Aneuploidy and LOH, which are hallmarks of cancer genomes, were mapped in HeLa by constructing a digital copy-number profile at kilobase resolution (Fig. 1, Supplementary Fig. 4 and Supplementary Table 7). Read coverage profiles were segmented by a Hidden Markov Model (HMM) and recalibrated to account for widespread aneuploidy (Supplementary Figs 5 and 6). Sixty-one per cent of the genome has a baseline copy number of three, and only a small minority (3%) has a copy number of greater than four or less than two (Supplementary Table 8). LOH encompassed 15.7% of the genome, including several entire chromosome arms (5p, 6q, Xp, Xq) or large distal portions (2q, 3q, 6p, 11q, 13q, 19p, 22q) (Supplementary Fig. 7 and Supplementary Table 9), consistent with previous descriptions of LOH in cervical carcinomas[17]. The overall profile is consistent with published karyotypes of various HeLa strains[4], suggesting that the hypertriploid state arose either during tumorigenesis or early in the establishment of the HeLa cell line.

[1]Department of Genome Sciences, University of Washington, Seattle, Washington 98115, USA.
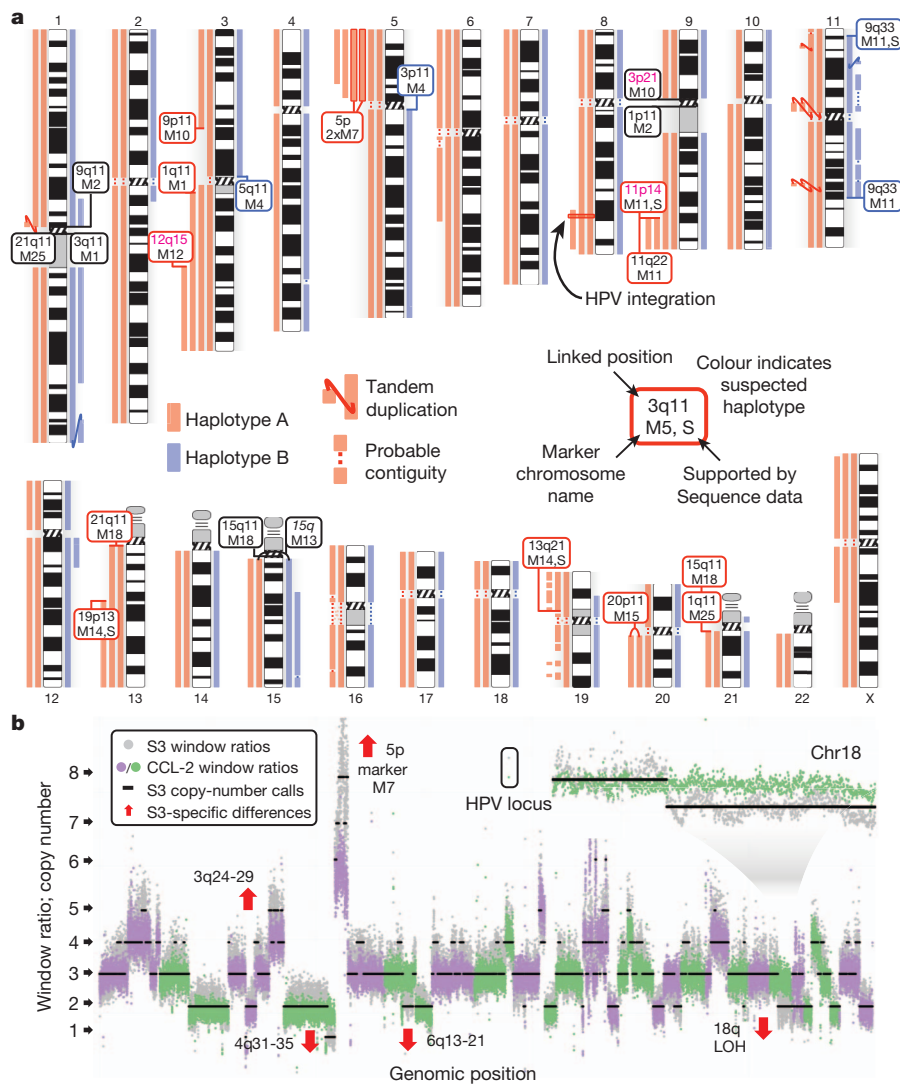*These authors contributed equally to this work.

Figure 1 | Haplotype-resolved copy number of the HeLa cancer cell line genome. a, Copy-number profile of HeLa split by haplotypes. Links denote likely contiguity and tandem duplications. Boxes indicate marker chromosomes identified by copy-number breakpoints (boxes are coloured by haplotype; black, unknown; pink text, uncertain locations; S, links confirmed by mate-pair sequencing). **b**, Windowed copy-number ratios for HeLa CCL-2 (green and purple, alternating chromosomes) and HeLa S3 (grey), with predicted integer copy number for S3 (black). Notable strain differences are indicated by red arrows (for example, reduced copy over chromosome 18q). The window containing the HPV insertion and rearrangement is at elevated copy in both strains.

Structural variants were identified by clustering discordantly mapped reads from 40-kb and 3-kb mate-pair libraries (Supplementary Fig. 8). Twenty interchromosomal links were identified, including links for marker chromosomes M11 (9q33–11p14) and M14 (13q21–19p13). In addition, 209 HeLa-specific deletions and 8 inversions were found (Supplementary Figs 9 and 11, and Supplementary Table 10). Only two genes that are impacted by HeLa-specific structural rearrangements (Supplementary Table 11) intersected with SCGC (*STK11* (ref. 18), *FHIT*), both of which are recurrently deleted in cervical carcinomas[18,19].

Conventional whole-genome sequencing fails to resolve haplotype phase, an essential aspect of the description and interpretation of non-haploid genomes, including cancer genomes[20]. Recently, several groups have demonstrated genome-wide measurement of local[5] or sparse[21] haplotypes, but these approaches have yet to be applied to aneuploid cancer genomes. To resolve haplotype phase across the HeLa genome, we sequenced pools of fosmid clones[5]. Specifically, we constructed three complex fosmid-clone libraries, and then carried out limiting dilution and shotgun sequencing of 288 fosmid clone pools. In summary, these were estimated to include 518,293 individual non-overlapping clones with a median insert size of 33 kb, for a total physical coverage of 6.3× of the haploid reference genome (Supplementary Fig. 12). The complement of likely inherited heterozygous variants (SNP and indel, $n = 1.97 \times 10^6$) was ascertained by shotgun sequencing and by cross-referencing with calls made by the 1000 Genomes Project, and then re-genotyped using reads from each clone

pool. Alleles that were present at distinct heterozygous sites within a given clone were assigned, or 'phased', to the same inherited haplotype, and the unobserved alleles were implicitly phased to the opposite haplotype. When overlapping clones from distinct pools were merged, this resulted in haplotype blocks with an N50 (the contig size above which 50% of the total length of the haplotype assembly is included) of 550 kb containing 90.6% of heterozygous variants that were probably inherited.

Most of the HeLa genome is present at an uneven haplotype ratio (for example, 2:1 in regions in which copy number = 3). We sought to exploit the resulting allelic imbalance to phase consecutive haplotype blocks (Supplementary Fig. 13). We first calculated the cumulative allelic ratio among shotgun reads for the SNVs residing in each haplotype block, which clustered closely with the underlying haplotype ratio. For example, in non-LOH regions with a copy number of 3 that have ratios of 2:1 or 1:2, allelic ratios calculated for each block had distributions centred on 0.32 or 0.65, close to the expected fractions of one-third and two-thirds (Supplementary Fig. 14). Using these ratios, we merged haplotype blocks into scaffolds covering 1.96 Gb or 90.3% of the non-LOH HeLa genome (scaffold N50 of 44.8 megabases (Mb); Supplementary Table 12). The haplotype-resolved scaffolds were then merged with the copy-number map to produce a global, haplotype-resolved copy-number profile of the aneuploid HeLa genome (Fig. 1a, Supplementary Fig. 15 and Supplementary Table 13).

Phasing accuracy was independently confirmed by several methods. First, 99.7% of informative read pairs from 3-kb mate-pair sequencing

(each read overlapping a phased site) were concordant with the predicted phase. Second, long-insert single-molecule sequencing (Pacific Biosciences RS; mean, 2.97 kb; 90th percentile, 5.1 kb among informative reads) showed that 97.2% of reads were in perfect agreement with the predicted phase, despite the high per-base sequencing error rate of approximately 15% (Supplementary Fig. 16). Third, examination of allelic state across 47.3 Mb of chromosome 18q, which underwent LOH in HeLa S3 but not in CCL-2, showed that out of the 17,761 affected alleles (heterozygous in CCL-2 but at an allele balance of greater than 0.9 among S3 reads), 99.7% corresponded to those phased together on haplotype A in CCL-2 (Supplementary Fig. 17). Finally, windowed analysis of population allele frequencies revealed probable African or European genetic ancestry across long stretches of the haplotype-resolved genome, consistent with recent admixture and a low switch error rate (Supplementary Figs 18 and 19).

To measure the frequency of new mutations in the HeLa genome, we examined amplified haplotypes for de facto somatic mutations occurring during tumorigenesis or early in the cell line's subsequent passaging. Within LOH regions, these appear as polymorphisms; 2,883 such sites (mean, 1.31 per haploid Mb; Supplementary Table 14) were confirmed by clone-pool sequencing and allele frequency in shotgun sequencing (Supplementary Figs 20 and 21). In non-LOH regions, in which one haplotype is amplified but both remain present, the majority of observed heterozygous sites are inherited, as reflected by their substantial overlap with variants from the 1000 Genomes Project (86.7%, $n = 2,339,608$). Excluding these and sites found in the 11 control genomes, 5,282 sites (mean, 1.32 per haploid Mb) remained at which clones differed in genotype between the two or more amplified copies of the same germline haplotype, with little regional variation in the abundance (Supplementary Fig. 22). In summary, 8,165 somatic mutations were validated with an estimated sensitivity of 61.1%, placing an upper bound on the point-mutational burden sustained by HeLa CCL-2 after aneuploidy. Despite many additional doublings in culture, this point-mutation frequency (2.16 per Mb) is on the lower end of frequencies observed across different cancer genomes[22]. However, without estimates for parameters such as the number of doublings during tumorigenesis, the count of cells explanted, and the number of passages in culture, this estimate of post-aneuploidy mutational burden cannot be rescaled to a rate per base per division.

Four years after the initial establishment of the HeLa cell line, several additional strains were cloned[23]. One of these, HeLa S3, remains in widespread use today and has been profiled extensively as part of the ENCODE Project. To investigate the divergence between CCL-2 and S3, we carried out shotgun sequencing of S3 to 26× coverage. Outside of S3-specific regions of LOH, 94.5% of rare variants in CCL-2 were shared with S3 ($n = 204,841$ sites excluding 1000 Genomes Project and segmental duplications, and requiring ≥8× coverage in each genome; Supplementary Fig. 23 and Supplementary Table 15). Somatic mutations were also shared, though to a lesser degree: 72.4% of clone-confirmed somatic mutations from CCL-2 were found in S3 ($n = 8,054$ sites with ≥8× coverage in S3), consistent with a low rate of somatic SNV accumulation since the strains diverged in 1955.

The copy-number profile of HeLa S3 broadly mirrors that of CCL-2 (Fig. 1b, and Supplementary Figs 7 and 24) as well as eight additional HeLa strains that we sequenced lightly (3.5 to 4.3×). We observed some strain-specific differences (Supplementary Figs 25–27), consistent with previous reports of karyotypic heterogeneity both among and within strains. Despite some variability, a copy number of three was the dominant state consistently, with a median of 52% of the genome across the eight strains (range 38–60%), similar to its prevalence in CCL-2 (61%). Gains or losses of entire chromosome arms were observed (for example, chr18q, HeLa S3 (Fig. 1b), chr9p, CCL-13 (Supplementary Figs 28 and 29)), but smaller amplifications and deletions were more common. These may correspond to variability in copy rather than in the content of marker chromosomes present, as suggested by high overall breakpoint concordance between strains (81% of

copy-number breakpoints within ±1 Mb were present in ≥2 strains). The additional eight cell lines analysed here were identified in the 1970s[24] as products of HeLa contamination into other tissue cultures in the preceding two decades. Their shared set of structural abnormalities reflects their common origin from small founder populations of contaminating cells and reinforces the view that the structural rearrangements resulting in marker chromosomes arose early and are variable in copy number.

Nearly all cervical cancer is caused by human papillomavirus (HPV) infection. Within HeLa, a partial copy of the HPV-18 genome is integrated at a known fragile site on chromosome 8q24.21 (refs 25, 26). Haplotype and copy-number maps indicate that the flanking regions are present at copy number four, at a haplotype ratio of 3:1. To characterize the structure and copy number of the insertion, we included the HPV-18 genome alongside the human reference during alignment of clone-pool reads. By analysing patterns of coverage from breakpoint-spanning fosmid clones, read-depth data and breakpoint sequencing, we generated a structural model for the viral integration (Fig. 2a, b, and Supplementary Figs 30 and 31). Two repeat structures (which we designate R1 and R2) consisting of the partial viral genome are interspersed with regions of human chromosome 8q24.21 genomic DNA. The viral genome is present with identical breakpoints on each copy of the amplified haplotype, to the exclusion of the other haplotype, which remains at single copy and lacks integration-associated rearrangements, confirming that integration and rearrangement preceded aneuploidy. The integrated structure contains only two-thirds of the complete HPV-18 genome, including full-length copies of the E6 and E7 oncogenes necessary for telomerase activity (amplified to a copy number of approximately 12), but lacking a functional copy of E2, an inhibitor of E6 and E7 (ref. 13) (Fig. 2c). In addition, a distinct portion of the HPV-18 genome, amplified to a copy number of approximately 30 in HeLa, includes an epithelium-specific enhancer
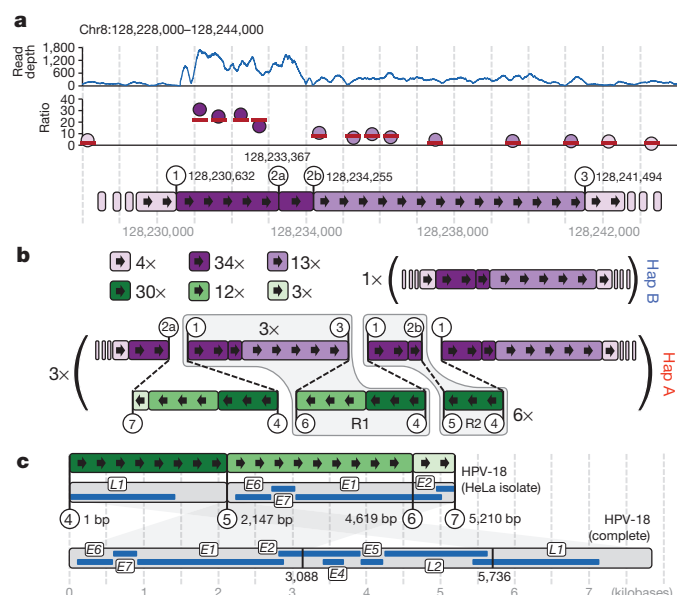

Figure 2 | HeLa HPV integration locus. a, Chromosome 8 read depth flanking the HPV integration site (top, blue line), windowed copy-number ratios (purple points, shaded by segment) and integer copy states (red bars, middle), and corresponding segments and breakpoints (circled numbers with genomic coordinates, bottom). **b,** Proposed HPV integration structure: per-segment copy number (top left), non-rearranged haplotype B (copy number = 1, top right), rearranged haplotype A with HPV insertion (copy number = 3, bottom) carrying approximately 3 and 6 tandem copies of repeats R1 and R2, respectively. Hap, haplotype. **c,** The partial HPV-18 genome and corresponding genes (grey and blue, top) with breakpoints highlighted by numbered circles. For reference, the entire HPV-18 genome is shown (bottom).
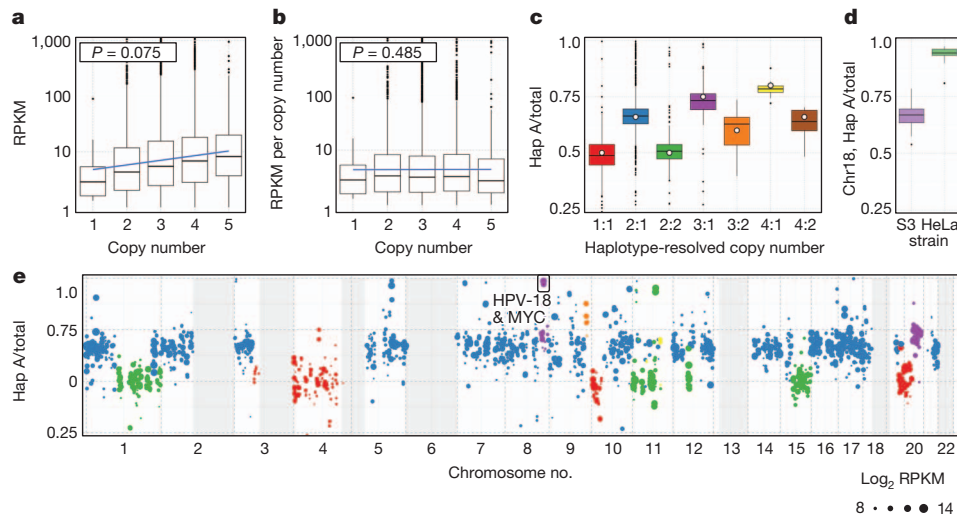
Figure 3 | Gene expression by copy number and haplotype in HeLa S3.
a, Transcript abundance (reads per kilobase per million (RPKM), for genes with an RPKM ≥1) is positively correlated with gene copy. **b**, Expression per copy (RPKM per gene copy number) does not correlate with copy number.
c, Fractional contribution of haplotype A to overall expression (Hap A/total) (RPKM averaged across megabase windows at phased sites) split by haplotype-resolved copy number. Open circles indicate expected fractions.
d, Haplotype-A-specific expression in HeLa S3 but not CCL-2 across S3-specific LOH on chr18q. **e**, Haplotype A fractional contribution to expression across the genome, colour-coded by underlying haplotype-resolved copy number as in **c** (point size represents the log₂ total RPKM, grey boxes indicate HeLa S3 LOH).

that controls *E6* and *E7* transcription[27], possibly contributing to their high expression (Supplementary Fig. 32).

Extensive sequencing-based functional genomic data have been generated on HeLa and other cancer cell lines by the ENCODE Project[7], but these have the potential to be misinterpreted if their analysis does not account for aneuploidy and phase. As HeLa CCL-2 and S3 are nearly identical in genotype, we used haplotype and copy-number maps of CCL-2 to assign phase to publicly available functional data generated on S3 (ref. 7), including transcription-factor binding, chromatin modification and chromatin-accessibility data sets. We also calculated haplotype-specific gene-expression scores using RNA sequencing (RNA-seq) data generated in this study and by others[6,7] (Supplementary Figs 33–35). For each data set, aligned reads were phased by comparison to HeLa CCL-2 haplotype blocks. Corresponding peak scores (chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) and DNase I sequencing (DNase-seq)) or gene-expression values (RNA-Seq) called from the full set of reads were divided proportionally based on the abundance of phase-informative mapping to each haplotype, normalized to each haplotype's estimated copy number. Mapping to the human reference genome imposed a slight bias, favouring the reference allele by an

average of 1.08-fold. We constructed two HeLa-specific reference sequences by introducing all SNVs from each haplotype onto one or the other; mapping to this reference mitigated most of the bias (to 1.02-fold, or a 75% reduction; Supplementary Figs 36–38).

Across the HeLa genome, gene expression is significantly correlated with copy number ($P = 0.075$; Fig. 3a, b), suggesting a minimal role for gene-dosage buffering. Moreover, on average, each haplotype copy makes a comparable contribution to the transcriptome, despite uneven amplification and, in some cases, rearrangement (Fig. 3c, e). This trend is also observed for histone modifications, DNase hypersensitivity and transcription factor binding (Supplementary Figs 39 and 40). Transcript allele balances at sites heterozygous in CCL-2 on chromosome 18q closely followed the genomic balance (mean 66% representation of the A allele (two-thirds was expected)), but S3 nearly exclusively matched the A allele (94% of reads), reflecting the S3-specific LOH event (Fig. 3d). However, a small number of regions showed strong imbalances between each haplotype's contribution to overall patterns of expression, chromatin modification and transcription-factor binding (2.4% of ENCODE peaks, excluding those in LOH regions; Supplementary Figs 41–44). Interestingly, the HPV-18 insertion locus and proto-oncogene *MYC* (separated by approximately
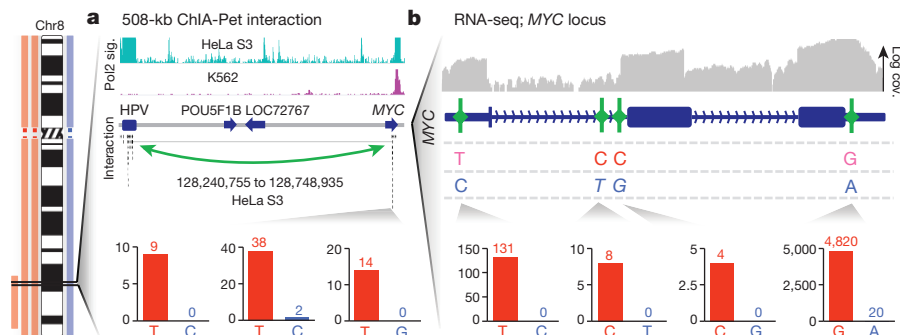
Figure 4 | Haplotype-specific regulation near the HPV integration site.
a, Long-range chromatin interactions between the HPV and *MYC* loci demonstrated by ChIA-PET[28] with the RNA polymerase II signal (top) shown for HeLa S3 and an HPV⁻ cell line (K562). Chromatin interactions (middle) are indicated by a green arrow. Bar graphs (bottom) show read counts at

phased, informative sites in *MYC* (red, haplotype A, blue, haplotype B).
b, Transcript abundance in HeLa S3 across the *MYC* locus measured by RNA-seq. Overall coverage is shown in grey (top) with phased, informative sites highlighted by green ticks (pink text, non-reference alleles). Haplotype contributions at each variant are shown in bar graphs (bottom), as in **a**.

500 kb) were among the regions with the most highly haplotype-imbalanced regulation in the genome (Supplementary Fig. 45). Phased RNA-seq data indicate that *MYC* is highly expressed, but almost exclusively from the HPV-18-integrated haplotype (mean ratio, 95:1; Fig. 4b and Supplementary Fig. 46). Phased ENCODE tracks and long-range chromatin interaction data (ChIA-PET (chromatin interaction analysis with paired-end tag sequencing)[28]; Fig. 4a and Supplementary Fig. 47) across the region indicate that transcription-factor occupancy, active chromatin marks and long-distance physical contacts are also nearly exclusive to the HPV-integrated, transcriptionally active haplotype. Taken together, these data implicate viral integration as a strong activator of *MYC* expression[29], acting in *cis* rather than in *trans* and possibly mediated by the epithelium-specific viral enhancer amplified to a copy number of approximately 30 within the R1 repeat structure (Fig. 2b)[27]. This strong *cis* interaction—between the amplified, integrated genome of a DNA tumour virus and a canonical proto-oncogene—may underlie the robust growth characteristics of the HeLa cell line, and provides indirect support for the hypothesis that inherited risk loci for cancer at chromosome 8q24 operate through activation of *MYC*[30].

In summary, we present a haplotype-resolved genome and a haplotype-resolved epigenome of a human cancer. Our study not only provides an overdue genomic analysis of the human cell line that is possibly the most commonly used in biomedical research but also represents a unique view into a cancer genome and epigenome enabled by the acquisition of haplotype information.

METHODS SUMMARY

Cells were maintained at 37 °C in DMEM F-12. Shotgun libraries were prepared by conventional ligation-based methods, sequenced on an Illumina HiSeq 2000 instrument. Point variants were called using shotgun sequence reads. Copy-number maps were created from read depth. Long-insert clone-dilution pools were created and analysed as described previously[5]. Data sets used for each analysis are depicted as a flow chart in Supplementary Fig. 48. Full methods and associated references can be found in the online version of the paper and in Supplementary Notes 1–23.

Full Methods and any associated references are available in the online version of the paper.

1. Gey, G. O., Coffman, W. D. & Kubicek, M. T. Tissue culture studies of the proliferative capacity of cervical carcinoma and normal epithelium. *Cancer Res.* **12,** 264–265 (1952).
2. Gartler, S. M. Apparent Hela cell contamination of human heteroploid cell lines. *Nature* **217,** 750–751 (1968).
3. Skloot, R. *The Immortal Life of Henrietta Lacks*. (Crown Publishers, 2010).
4. Macville, M. *et al.* Comprehensive and definitive molecular cytogenetic characterization of HeLa cells by spectral karyotyping. *Cancer Res.* **59,** 141–150 (1999).
5. Kitzman, J. O. *et al.* Haplotype-resolved genome sequencing of a Gujarati Indian individual. *Nature Biotechnol.* **29,** 59–63 (2011).
6. Nagaraj, N. *et al.* Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* **7,** 548 (2011).
7. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
8. Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338,** 222–226 (2012).
9. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).
10. Exome Variant Server. http://evs.gs.washington.edu/EVS/ (NHLBI GO Exome Sequencing Project (ESP), January 2012).
11. Morin, R. *et al.* Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *Biotechniques* **45,** 81–94 (2008).
12. The Cancer Genome Project. http://www.sanger.ac.uk/genetics/CGP/ (Wellcome Trust Sanger Institute, January 2013).
13. Goodwin, E. C. *et al.* Rapid induction of senescence in human cervical carcinoma cells. *Proc. Natl Acad. Sci. USA* **97,** 10978–10983 (2000).
14. Rosty, C. *et al.* Clinical and biological characteristics of cervical neoplasias with FGFR3 mutation. *Mol. Cancer* **4,** 15 (2005).
15. Talora, C., Sgroi, D. C., Crum, C. P. & Dotto, G. P. Specific down-modulation of Notch1 signaling in cervical cancer cells is required for sustained HPV-E6/E7 expression and late steps of malignant transformation. *Genes Dev.* **16,** 2252–2263 (2002).
16. White, E. A. *et al.* Comprehensive analysis of host cellular interactions with human papillomavirus E6 proteins identifies new E6 binding partners and reflects viral diversity. *J. Virol.* **86,** 13174–13186 (2012).
17. Corver, W. E. *et al.* Genome-wide allelic state analysis on flow-sorted tumor fractions provides an accurate measure of chromosomal aberrations. *Cancer Res.* **68,** 10333–10340 (2008).
18. Wingo, S. N. *et al.* Somatic LKB1 mutations promote cervical cancer progression. *PLoS ONE* **4,** e5137 (2009).
19. Wistuba, I. I. *et al.* Deletions of chromosome 3p are frequent and early events in the pathogenesis of uterine cervical carcinoma. *Cancer Res.* **57,** 3154–3158 (1997).
20. Nik-Zainal, S. *et al.* The life history of 21 breast cancers. *Cell* **149,** 994–1007 (2012).
21. Fan, H. C., Wang, J., Potanina, A. & Quake, S. R. Whole-genome molecular haplotyping of single cells. *Nature Biotechnol.* **29,** 51–57 (2011).
22. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489,** 519–525 (2012); corrigendum **491,** 288 (2012).
23. Puck, T. T. & Marcus, P. I. A rapid method for viable cell titration and clone production with HeLa cells in tissue culture: the use of X-irradiated cells to supply conditioning factors. *Proc. Natl Acad. Sci. USA* **41,** 432–437 (1955).
24. Nelson-Rees, W. A., Daniels, D. W. & Flandermeyer, R. R. Cross-contamination of cells in culture. *Science* **212,** 446–452 (1981).
25. Wentzensen, N., Vinokurova, S. & von Knebel Doeberitz, M. Systematic review of genomic integration sites of human papillomavirus genomes in epithelial dysplasia and invasive cancer of the female lower genital tract. *Cancer Res.* **64,** 3878–3884 (2004).
26. Lazo, P. A., DiPaolo, J. A. & Popescu, N. C. Amplification of the integrated viral transforming genes of human papillomavirus 18 and its 5′-flanking cellular sequence located near the *myc* protooncogene in HeLa cells. *Cancer Res.* **49,** 4305–4310 (1989).
27. Bouallaga, I., Massicard, S., Yaniv, M. & Thierry, F. An enhanceosome containing the Jun B/Fra-2 heterodimer and the HMG-I(Y) architectural protein controls HPV 18 transcription. *EMBO Rep.* **1,** 422–427 (2000).
28. Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148,** 84–98 (2012).
29. Peter, M. *et al.* MYC activation associated with the integration of HPV DNA at the MYC locus in genital tumors. *Oncogene* **25,** 5985–5993 (2006).
30. Ahmadiyeh, N. *et al.* 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with *MYC*. *Proc. Natl Acad. Sci. USA* **107,** 9742–9746 (2010).

Supplementary Information is available in the online version of the paper.

METHODS

HeLa cell culture. HeLa cell cultures (HeLa ATCC, CCL-2 (laboratory stock); HeLa S3 ATCC, CCL-2.2 (laboratory stock); Chang liver ATCC, CCL-13; L132 ATCC, CCL-5; KB ATCC, CCL-17; HEp-2 ATCC, CCL-23; WISH ATCC, CCL-25; Intestine 407 ATCC, CCL-6; FL ATCC, CCL-62; AV-3 ATCC, CCL-21) were maintained in DMEM F-12, HEPES (Gibco) media supplemented with fetal bovine serum (FBS) to 10% and a $1\times$ final concentration of pen-strep antibiotic (Gibco).

Shotgun sequencing, alignment and variant calling. All shotgun libraries were constructed using standard ligation chemistry methods and sequenced on an Illumina HiSeq 2000. Reads were aligned to the human reference genome (hg19, b37) using BWA[31] followed by duplicate removal, quality score recalibration and local indel realignment using GATK[32]. SNVs were called using samtools[33], indel variants were called using GATK[32] and short tandem repeats (STRs) were called using LobSTR[34] (Supplementary Note 1). Indel detection as a function of coverage was investigated further as described in Supplementary Note 2. Gene ontology term analysis was carried out using DAVID[35]. Data sets used for each analysis are depicted as a flow chart in Supplementary Fig. 48.

Read depth copy number analysis. Shotgun reads for HeLa and Human Genome Diversity Project (HGDP) control genomes[8] along with a similarly prepared control library with a matched G + C profile were aligned using mrsFAST[36], processed as described previously[37] to generate read depth-based copy number predictions within non-overlapping windows of singly unique nucleotide k-mers (SUNK windows; Supplementary Note 3). Copy-number calling in HeLa was carried out at high (approximately 1.5-kb) and low (approximately 77-kb) resolution using an HMM (Supplementary Note 4), and a recalibration process was then used to account for widespread aneuploidy (Supplementary Note 5). Short amplifications and deletions were identified using a sliding-window approach (Supplementary Note 6). Copy-number calling was also carried out on HeLa S3 at both high and low resolutions, as well as on the eight additional HeLa strains at low resolution, and profiles were compared between strains (Supplementary Note 7). Regions of LOH were identified using a two-state HMM that used the fraction of homozygous SNVs in non-repetitive regions across low-resolution copy-number windows described above (Supplementary Note 8).

Mate-pair library construction, sequencing and analysis. Library construction for 40-kb mate-pair libraries was carried out starting with fosmid clone DNA pooled within each original fosmid preparation, using a protocol similar to one described previously[38] (Supplementary Note 9). Libraries of approximately 3-kb inserts were constructed following protocols described previously[39] (Supplementary Note 9). After read trimming and alignment, reads were split into classes based on aligned orientation and insert size, and processed using sliding windows to identify regions of probable structural rearrangement (Supplementary Note 10).

Fosmid pool construction, sequencing and haplotype phasing. Three replicate fosmid libraries were prepared as described previously[5], and then partitioned by limited dilution into 96 sub-libraries. This was followed by outgrowth, barcoded transposase-based library preparation[40], sequencing and alignment (Supplementary Note 11). Clone boundaries were inferred as described previously[5], and base calls were made at all heterozygous variant positions as ascertained from whole-genome shotgun sequencing. Overlapping clones were merged to consensus haplotype blocks using an implementation of the ReFHap algorithm[41] (Supplementary Note 12). Within the majority of the HeLa genome in which haplotypes are unequally amplified, adjacent blocks were merged to create scaffolds, using an HMM that finds the most likely phase of neighbouring blocks given their shotgun allele frequencies of inherited variants (those found within the 1000 Genomes Project, Supplementary Note 12). This produced a final set of haplotype scaffolds with an N50 size of 44.8 Mb, which was then used in conjunction with copy-number calls to estimate haplotype-resolved copy number for HeLa (Supplementary Note 13). Haplotype scaffolds were analysed for variant population frequencies to investigate the ancestral origin of phased blocks (Supplementary Note 14). Finally, overall copy numbers were compared among all HeLa strains sequenced in this study (Supplementary Note 15).

Long-read phase validation. Genomic DNA from HeLa CCL-2 was mechanically sheared using a Covaris G-tube column and standard microcentrifuge following the manufacturer's instructions, and this produced a mean fragment size of approximately 10 kb. Single-molecule real-time sequencing libraries for the Pacific Biosciences RS sequencer were prepared using the Pacific Biosciences DNA Template Prep Kit (3–10 kb), and the resulting library was sequenced across eight cells using a 90-min movie. Resulting base calls were aligned to the genome with bwasw (using parameters '-b5 -q2 -r1 -z1'). Reads that overlapped at least two phased SNPs were considered, excluding those within ±10 bp of an insertion or deletion in the alignment.

Identification of putative post-aneuploidy mutations. We searched for candidate somatic post-aneuploidy mutations by taking the initial set of SNVs called from the shotgun sequencing data and filtering to remove probable germline variants. SNVs that were phased on a duplicated haplotype but that were polymorphic between the two duplicated copies were identified. Common polymorphisms and sequencing artefacts were removed by filtering against repeat annotations and control genomes (Supplementary Note 16).

HPV-18 insertion characterization. The HPV-18 integration locus was characterized by aligning all fosmid libraries to a modified genome that included the HPV-18 reference genome as an additional chromosome. Interchromosomal read pairs, fosmid-pool coverage profiles, and copy-number calls were used to determine the repeat structure of the chromosome 8q24.21–HPV-18 integration locus. Polymerase-chain-reaction primers were then designed to amplify the proposed breakpoints, and then sequencing for base-pair resolution was carried out (Supplementary Note 17).

ENCODE and RNA-seq phasing. Directional, PolyA$^+$ RNA-seq data generated in-house on HeLa S3 (Supplementary Note 18) were analysed in parallel with publically available ENCODE epigenomics and transcriptomics data downloaded from the online data portal for HeLa S3, and RNA-seq data on HeLa CCL-2 (ref. 6) (Supplementary Note 19). RNA-seq reads were aligned using TopHat[42] and transcript quantification was carried out using Cufflinks[43]. Haplotype phasing was performed by genotyping aligned-sequence data for all phased SNVs and assigning haplotype contributions to either peaks (epigenomics data sets) or RPKM (RNA-seq data sets), and then carrying out copy-number normalization (Supplementary Note 20). Reference bias was investigated in all tracks and removed in a subset to identify its impact on outlier calling (Supplementary Note 21). Haplotype-specific peaks were then identified in all data tracks (Supplementary Note 22). Finally, a meta-analysis of all data tracks was used to identify large regions of haplotype imbalance (Supplementary Note 23).

31. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
32. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–1303 (2010).
33. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27,** 2987–2993 (2011).
34. Gymrek, M., Golan, D., Rosset, S. & Erlich, Y. lobSTR: a short tandem repeat profiler for personal genomes. *Genome Res.* **22,** 1154–1162 (2012).
35. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4,** 44–57 (2009).
36. Hach, F. *et al.* mrsFAST: a cache-oblivious algorithm for short-read mapping. *Nature Methods* **7,** 576–577 (2010).
37. Sudmant, P. H. *et al.* Diversity of human copy number variation and multicopy genes. *Science* **330,** 641–646 (2010).
38. Gnerre, S. *et al.* High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl Acad. Sci. USA* **108,** 1513–1518 (2011).
39. Talkowski, M. E. *et al.* Next-generation sequencing strategies enable routine detection of balanced chromosome rearrangements for clinical diagnostics and genetic research. *Am. J. Hum. Genet.* **88,** 469–481 (2011).
40. Adey, A. *et al.* Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome. Biol.* **11,** R119 (2010).
41. Duitama, J. *et al.* Fosmid-based whole genome haplotyping of a HapMap trio child: evaluation of single individual haplotyping techniques. *Nucleic Acids Res.* **40,** 2041–2053 (2012).
42. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25,** 1105–1111 (2009).
43. Roberts, A., Pimentel, H., Trapnell, C. & Pachter, L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* **27,** 2325–2329 (2011).

LETTER

A directional tuning map of *Drosophila* elementary motion detectors

Matthew S. Maisak[1]*, Juergen Haag[1]*, Georg Ammer[1], Etienne Serbe[1], Matthias Meier[1], Aljoscha Leonhardt[1], Tabea Schilling[1], Armin Bahl[1], Gerald M. Rubin[2], Aljoscha Nern[2], Barry J. Dickson[3], Dierk F. Reiff[1]†, Elisabeth Hopp[1] & Alexander Borst[1]

The extraction of directional motion information from changing retinal images is one of the earliest and most important processing steps in any visual system. In the fly optic lobe, two parallel processing streams have been anatomically described, leading from two first-order interneurons, L1 and L2, via T4 and T5 cells onto large, wide-field motion-sensitive interneurons of the lobula plate[1]. Therefore, T4 and T5 cells are thought to have a pivotal role in motion processing; however, owing to their small size, it is difficult to obtain electrical recordings of T4 and T5 cells, leaving their visual response properties largely unknown. We circumvent this problem by means of optical recording from these cells in *Drosophila*, using the genetically encoded calcium indicator GCaMP5 (ref. 2). Here we find that specific subpopulations of T4 and T5 cells are directionally tuned to one of the four cardinal directions; that is, front-to-back, back-to-front, upwards and downwards. Depending on their preferred direction, T4 and T5 cells terminate in specific sublayers of the lobula plate. T4 and T5 functionally segregate with respect to contrast polarity: whereas T4 cells selectively respond to moving brightness increments (ON edges), T5 cells only respond to moving brightness decrements (OFF edges). When the output from T4 or T5 cells is blocked, the responses of postsynaptic lobula plate neurons to moving ON (T4 block) or OFF edges (T5 block) are selectively compromised. The same effects are seen in turning responses of tethered walking flies. Thus, starting with L1 and L2, the visual input is split into separate ON and OFF pathways, and motion along all four cardinal directions is computed separately within each pathway. The output of these eight different motion detectors is then sorted such that ON (T4) and OFF (T5) motion detectors with the same directional tuning converge in the same layer of the lobula plate, jointly providing the input to downstream circuits and motion-driven behaviours.

Most of the neurons in the fly brain are dedicated to image processing. The respective part of the head ganglion, called the optic lobe, consists of several layers of neuropile called lamina, medulla, lobula and lobula plate, all built from repetitive columns arranged in a retinotopic way (Fig. 1a). Each column houses a set of identified neurons that, on the basis of Golgi staining, have been described anatomically in great detail[3–5]. Owing to their small size, however, most of these columnar neurons have never been recorded from electrophysiologically. Therefore, their specific functional role in visual processing is still largely unknown. This fact is contrasted by rather detailed functional models about visual processing inferred from behavioural studies and recordings from the large, electrophysiologically accessible output neurons of the fly lobula plate (tangential cells). As the most prominent example of such models, the Reichardt detector derives directional motion information from primary sensory signals by multiplying the output from adjacent photoreceptors after asymmetric temporal filtering[6]. This model makes a number of rather counter-intuitive predictions all of which have been confirmed experimentally (for review, see ref. 7). Yet, the neurons corresponding to most
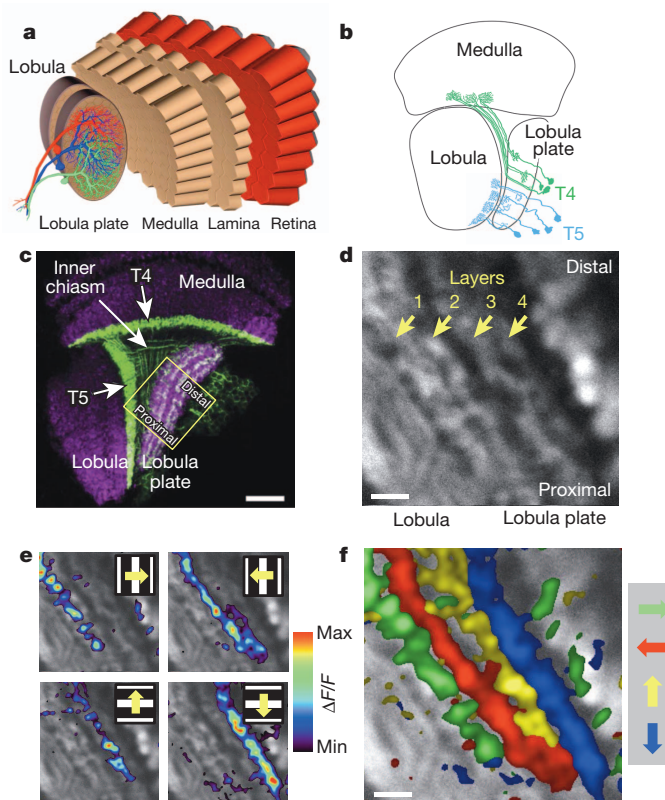

Figure 1 | Directional tuning and layer-specific projection of T4 and T5 cells. a, Schematic diagram of the fly optic lobe. In the lobula plate, motion-sensitive tangential cells extend their large dendrites over many hundreds of columns. Shown are the reconstructions of the three cells of the horizontal system[22]. **b**, Anatomy of T4 and T5 cells, as drawn from Golgi-impregnated material (from ref. 5). **c**, Confocal image of the Gal4-driver line R42F06, shown in a horizontal cross-section (from ref. 10). Neurons are marked in green (Kir2.1–EGFP labelled), whereas the neuropile is stained in purple by an antibody against the postsynaptic protein Dlg. Scale bar, 20 μm. **d**, Two-photon image of the lobula plate of a fly expressing GCaMP5 under the control of the same driver line R42F06. Scale bar, 5 μm. The size and orientation of the image approximately corresponds to the yellow square in **c. e**, Relative fluorescence changes (ΔF/F) obtained during 4-s grating motion along the four cardinal directions, overlaid on the greyscale image. Each motion direction leads to activity in a different layer. Minimum and maximum ΔF/F values were 0.3 and 1.0 (horizontal motion), and 0.15 and 0.6 (vertical motion). **f**, Compound representation of the results obtained from the same set of experiments. Scale bar, 5 μm. Results in **e** and **f** represent the data obtained from a single fly averaged over four stimulus repetitions. Similar results were obtained from six other flies.

[1]Max Planck Institute of Neurobiology, 82152 Martinsried, Germany. [2]Janelia Farm Research Campus, Ashburn, Virginia 20147, USA. [3]Institute of Molecular Pathology, 1030 Vienna, Austria. †Present address: Institute Biology 1, Albert-Ludwigs University, 79085 Freiburg, Germany.
*These authors contributed equally to this work.

of the circuit elements of the Reichardt detector have not been identified so far. Here, we focus on a set of neurons called T4 and T5 cells (Fig. 1b) which, on the basis of circumstantial evidence, have long been speculated to be involved in motion detection[1,8–10]. However, it is unclear to what extent T4 and T5 cells are directionally selective or whether direction selectivity is computed or enhanced within the dendrites of the tangential cells. Another important question concerns the functional separation between T4 and T5 cells; that is, whether they carry equivalent signals, maybe one being excitatory and the other inhibitory on the tangential cells, or whether they segregate into directional- and non-directional pathways[11] or into separate ON- and OFF-motion channels[12,13].

To answer these questions, we combined Gal4-driver lines specific for T4 and T5 cells[14] with GCaMP5 (ref. 2) and optically recorded the visual response properties using two-photon fluorescence microscopy[15]. In a first series of experiments, we used a driver line labelling both T4 and T5 cells. A confocal image (Fig. 1c, modified from ref. 10) revealed clear labelling (in green) in the medulla (T4 cell dendrites), in the lobula (T5 cell dendrites), as well as in four distinct layers of the lobula plate, representing the terminal arborizations of the four subpopulations of both T4 and T5 cells. These four layers of the lobula plate can also be seen in the two-photon microscope when the calcium indicator GCaMP5 is expressed (Fig. 1d). After stimulation of the fly with grating motion along four cardinal directions (front-to-back, back-to-front, upwards and downwards), activity is confined to mostly one of the four layers, depending on the direction in which the grating is moving (Fig. 1e). The outcome of all four stimulus conditions can be combined into a single image by assigning a particular colour to each pixel depending on the stimulus direction to which it responded most strongly (Fig. 1f). From these experiments it is clear that the four subpopulations of T4 and T5 cells produce selective calcium signals depending on the stimulus direction, in agreement with previous deoxyglucose labelling[8]. Sudden changes of the overall luminance evokes no responses in any of the layers (field flicker; $n = 4$ experiments, data not shown). However, gratings flickering in counter-phase lead to layer-specific responses, depending on the orientation of the grating (Supplementary Fig. 1).

The retinotopic arrangement of this input to the lobula plate is demonstrated by experiments where a dark edge was moved within a small area of the visual field only. Depending on the position of this area, activity of T4 and T5 cells is confined to different positions within the lobula plate (Fig. 2a). Consequently, when moving a bright vertical edge horizontally from back to front, activity of T4 and T5 cells is elicited sequentially in layer 2 of the lobula plate (Fig. 2b). These two experiments also demonstrate that T4 and T5 cells indeed signal motion locally. We next investigated the question of where direction selectivity of T4 and T5 cells arises; that is, whether it is already present in the dendrite, or whether it is generated by synaptic interactions within the lobula plate. This question is hard to answer, as the dendrites of both T4 and T5 cells form a dense mesh within the proximal layer of the medulla (T4) and the lobula (T5), respectively. However, signals within the inner chiasm where individual processes of T4 and T5 cells can be resolved in some preparations show a clear selectivity for motion in one over the other directions (Fig. 2c). Such signals are as directionally selective as the ones measured within the lobula plate, demonstrating that the signals delivered from the dendrites of T4 and T5 cells are already directionally selective.

To assess the particular contribution of T4 and T5 cells to the signals observed in the above experiments, we used driver lines specific for T4 and T5 cells, respectively. Applying the same stimulus protocol and data evaluation as in Fig. 1, identical results were obtained as before for both the T4- as well as the T5-specific driver line (Fig. 3a, b). We conclude that T4 and T5 cells each provide directionally selective signals to the lobula plate, in contrast to previous reports[11]. Thus, both T4 and T5 cells can be grouped, according to their preferred direction, into four subclasses covering all four cardinal directions, reminiscent of ON–OFF ganglion cells of the rabbit retina[16].
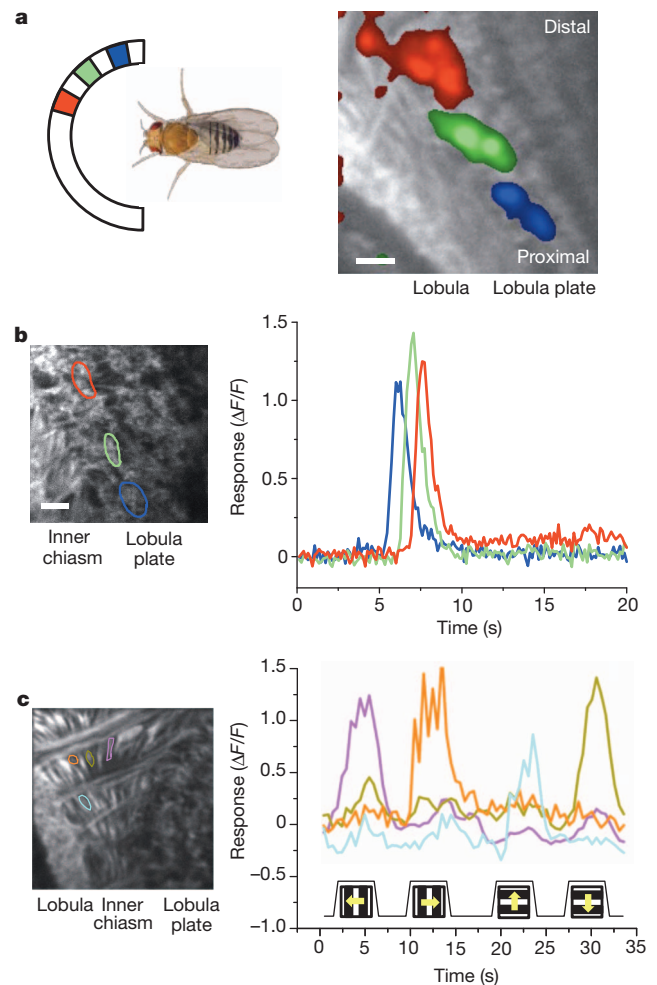
Figure 2 | Local signals of T4 and T5 cells. a, Retinotopic arrangement of T4 and T5 cells. A dark edge was moving repeatedly from front-to-back within a 15° wide area at different azimuthal positions (left). This leads to relative fluorescence changes at different positions along the proximal–distal axis within layer 1 of the lobula plate (right). Scale bar, 5 μm. Similar results have been obtained in four other flies. **b**, Sequential activation of T4 and T5 cells. A bright edge was moving from back-to-front at 15° s⁻¹. Scale bar, 5 μm. Similar results have been obtained in six other flies. **c**, Signals recorded from individual fibres within the inner chiasm (left) reveal a high degree of direction selectivity (right). Scale bar, 5 μm. Similar results were obtained from four other flies, including both lines specific for T4 and T5 cells. Response traces in **b** and **c** are derived from the region of interest encircled in the image with the same colour.

We next addressed whether T4 cells respond differently to T5 cells. To answer this question, we used, instead of gratings, moving edges with either positive (ON edge, brightness increment) or negative (OFF edge, brightness decrement) contrast polarity as visual stimuli. We found that T4 cells strongly responded to moving ON edges, but showed little or no response to moving OFF edges (Fig. 3c). This is true for T4 cells terminating in each of the four layers. We found the opposite for T5 cells. T5 cells selectively responded to moving OFF edges and mostly failed to respond to moving ON edges (Fig. 3d). Again, we found this for T5 cells in each of the four layers. We next addressed whether there are any other differences in the response properties between T4 and T5 cells by testing the velocity tuning of both cell populations by means of stimulating flies with grating motion along the horizontal axis from the front to the back at various velocities covering two orders of magnitude. T4 cells revealed a maximum response at a stimulus velocity of 30° s⁻¹, corresponding to a temporal frequency of 1 Hz (Fig. 3e). T5 cell responses showed a similar dependency on stimulus velocity, again with a peak at a temporal frequency of
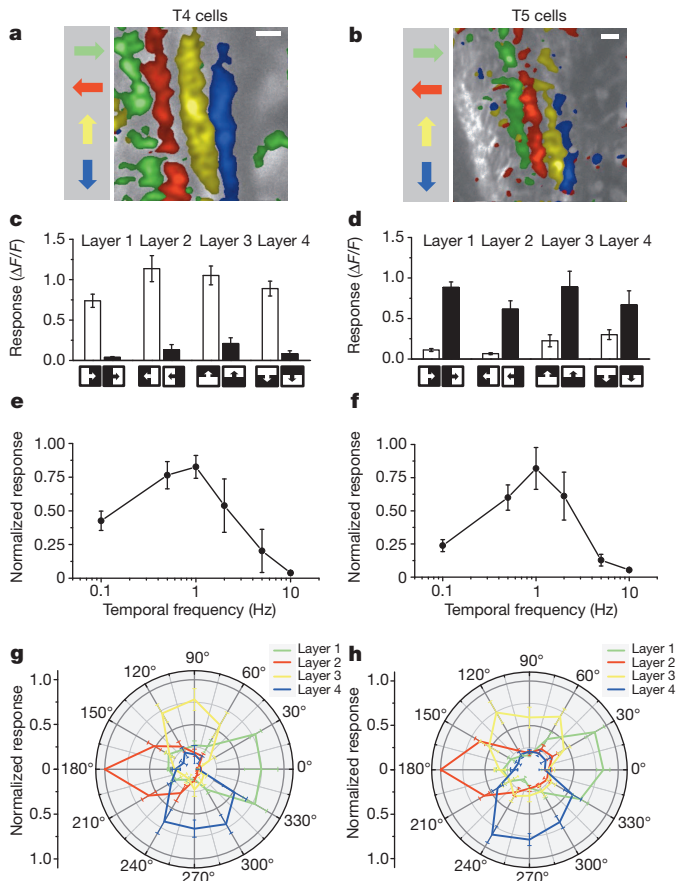
Figure 3 | Comparison of visual response properties between T4 and T5 cells. a, b, Relative fluorescence changes (Δ*F/F*) of the lobula plate terminals of T4 (**a**) and T5 (**b**) cells obtained during grating motion along the four cardinal directions. Results represent the data obtained from a single fly each, averaged over two stimulus repetitions. Scale bars, 5 μm. Similar results have been obtained in ten other flies. **c, d**, Responses of T4 (**c**) and T5 (**d**) cells to ON and OFF edges moving along all four cardinal directions. ON (white) and OFF (black) responses within each layer are significantly different from each other, with $P < 0.005$ except for layers 3 and 4 in T5 cells, where $P < 0.05$. **e, f**, Responses of T4 (**e**) and T5 (**f**) cells to gratings moving horizontally at different temporal frequencies. Relative fluorescence changes were evaluated from layer 1 of the lobula plate and normalized to the maximum response before averaging. **g, h**, Responses of T4 (**g**) and T5 (**h**) cells to gratings moving in 12 different directions. Relative fluorescence changes were evaluated from all four layers of the lobula plate normalized to the maximum response before averaging. Data represent the mean ± s.e.m. of the results obtained in $n = 8$ (**c**), $n = 7$ (**d**), $n = 6$ (**e**), $n = 7$ (**f**), $n = 6$ (**g**) and $n = 5$ (**h**) different flies. Significances indicated are based on two-sample *t*-test.

1 Hz (Fig. 3f). Thus, there is no obvious difference in the velocity tuning between T4 and T5 cells. As another possibility, T4 cells might functionally differ from T5 cells with respect to their directional tuning width. To test this, we stimulated flies with gratings moving into 12 different directions and evaluated the relative change of fluorescence in all four layers of the lobula plate. Using the T4-specific driver line, we found an approximate half width of 60–90° of the tuning curve, with the peak responses in each layer shifted by 90° (Fig. 3g). No decrease of calcium was detectable for grating motion opposite to the preferred direction of the respective layer. When we repeated the experiments using the T5-specific driver line, we found a similar dependence of the relative change of fluorescence on the stimulus direction (Fig. 3h). We conclude that T4 cells have the same velocity and orientation tuning as T5 cells. The only functional difference we were able to detect remains their selectivity for contrast polarity.

Our finding about the different preference of T4 and T5 cells for the polarity of a moving contrast makes the strong prediction that selective

blockade of T4 or T5 cells should selectively compromise the responses of downstream lobula plate tangential cells to either ON or OFF edges. To test this prediction, we blocked the output of either T4 or T5 cells via expression of the light chain of tetanus toxin[17] and recorded the responses of tangential cells via somatic whole-cell patch to moving ON and OFF edges. In response to moving ON edges, strong and reliable directional responses were observed in all control flies (Fig. 4a). However, T4-block flies showed a strongly reduced response to ON edges, whereas the responses of T5-block flies were at the level of control flies (Fig. 4b, c). When we used moving OFF edges, control flies again responded with a large amplitude (Fig. 4d). However, the responses of T4-block flies were at the level of control flies, whereas the responses of T5-block flies were strongly reduced (Fig. 4e, f). These findings are reminiscent on the phenotypes obtained from blocking lamina cells L1 and L2 (ref. 13) and demonstrate that T4 and T5 cells are indeed the motion-coding intermediaries for these contrast polarities on their way to the tangential cells of the lobula plate. Whether the residual responses to ON edges in T4-block flies and to OFF edges in T5-block flies are due to an incomplete signal separation between the two pathways or due to an incomplete genetic block in both fly lines is currently unclear.

To address the question of whether T4 and T5 cells are the only motion detectors of the fly visual system, or whether they represent one cell class, in parallel to other motion-sensitive elements, we used tethered flies walking on an air-suspended sphere[18] and stimulated them by ON and OFF edges moving in opposite directions[19]. As in the previous experiments, we blocked T4 and T5 cells specifically by selective expression of the light chain of tetanus toxin. During balanced motion, control flies did not show significant turning responses to either side (Fig. 4g). T4-block flies, however, strongly followed the direction of the moving OFF edges, whereas T5-block flies followed the direction of the moving ON edges (Fig. 4h, i). In summary, the selective preference of T4-block flies for OFF edges and of T5-block flies for ON edges not only corroborates our findings about the selective preference of T4 and T5 cells for different contrast polarities, but also demonstrates that the signals of T4 and T5 cells are indeed the major, if not exclusive, inputs to downstream circuits and motion-driven behaviours.

Almost a hundred years after T4 and T5 cells have been anatomically described[3], this study reports their functional properties in a systematic way. Using calcium as a proxy for membrane voltage[20], we found that both T4 and T5 cells respond to visual motion in a directionally selective manner and provide these signals to each of the four layers of the lobula plate, depending on their preferred direction. Both cell types show identical velocity and orientation tuning which matches the one of the tangential cells[21,22]. The strong direction selectivity of both T4 and T5 cells is unexpected, as previous studies had concluded that the high degree of direction selectivity of tangential cells is due to a push–pull configuration of weakly directional input with opposite preferred direction[23,24]. Furthermore, as the preferred direction of T4 and T5 cells matches the preferred direction of the tangential cells branching within corresponding layers, it is currently unclear which neurons are responsible for the null-direction response of the tangential cells. As for the functional separation between T4 and T5 cells, we found that T4 cells selectively respond to brightness increments, whereas T5 cells exclusively respond to moving brightness decrements. Interestingly, parallel ON and OFF motion pathways had been previously postulated on the basis of selective silencing of lamina neurons L1 and L2 (ref. 13). Studies using apparent motion stimuli to probe the underlying computational structure arrived at controversial conclusions: whereas some studies concluded that there was a separate handling of ON and OFF events by motion detectors[12,25,26], others did not favour such a strict separation[19,27]. The present study directly demonstrates the existence of separate ON and OFF motion detectors, as represented by T4 and T5 cells, respectively. Furthermore, our results anatomically confine the essential processing steps of elementary
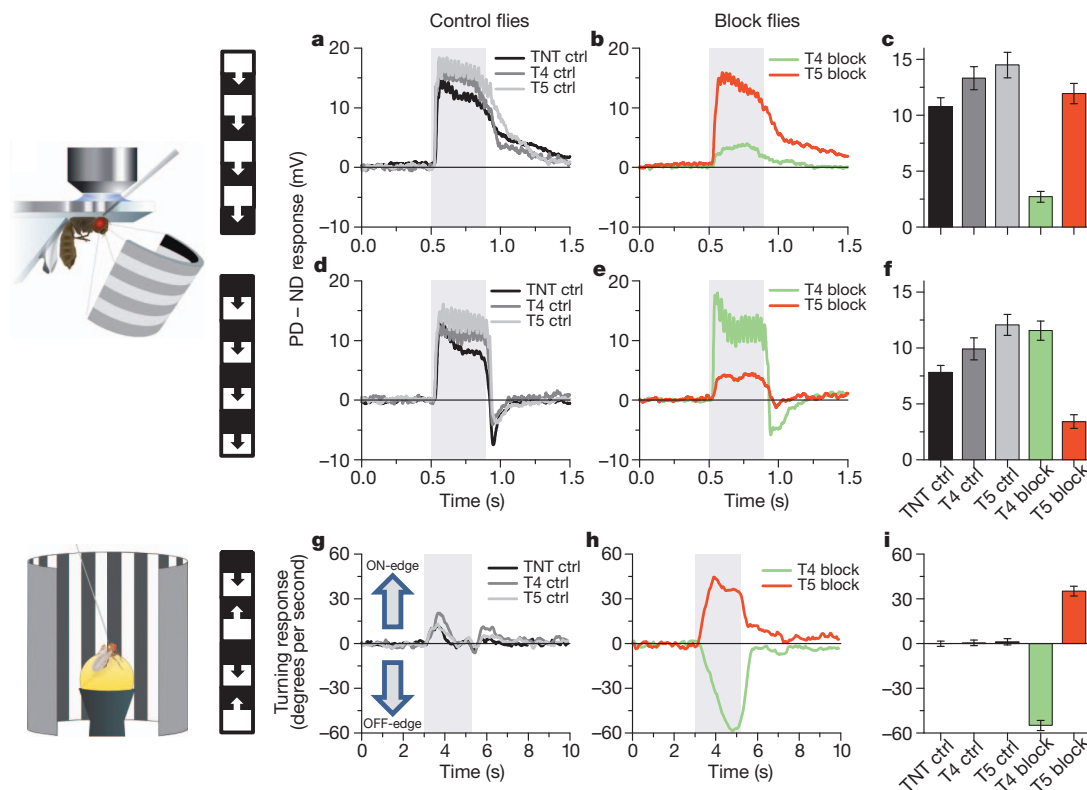
Figure 4 | Voltage responses of lobula plate tangential cells and turning responses of walking flies to moving ON and OFF edges. a, d, Average time course of the membrane potential in response to preferred direction motion minus the response to null direction motion (PD − ND response) as recorded in three types of control flies (stimulation period indicated by shaded area). **b, e**, Same as in **a, d**, but recorded in T4-block flies (green) and T5-block flies (red). The stimulus pattern, shown to the left, consisted of multiple ON- (**a**) or OFF-edges (**d**). **c, f**, Mean voltage responses (PD − ND) of tangential cells in the five groups of flies. Recordings were done from cells of the vertical[21] and the horizontal[22] system. Because no difference was detected between them, data were pooled. Data comprise recordings from n = 20 (TNT control), n = 12 (T4 control), n = 16 (T5 control), n = 17 (T4 block) and n = 18 (T5 block) cells. In both T4 and T5-block flies, ON and OFF responses are significantly different

from each other with P < 0.001. In T4-block flies, ON responses are significantly reduced compared to all three types of control flies, whereas in T5-block flies, OFF responses are significantly reduced, both with P < 0.001. **g**, Average time course of the turning response of three types of control flies to ON and OFF edges moving simultaneously to opposite directions (stimulation period indicated by shaded area). **h**, Same as in **g**, but recorded from T4-block flies (green) and T5-block flies (red). **i**, Mean turning tendency (±s.e.m.) during the last second of the stimulation period averaged across all flies within each group. Data comprise average values obtained in n = 12 (TNT controls), n = 11 (T4 controls), n = 11 (T5 controls), n = 13 (T4 block) and n = 12 (T5 block) flies. Values of T4 and T5-block flies are highly significantly different from zero with P < 0.001. Significances indicated are based on two-sample t-test.

motion detection—that is, asymmetric temporal filtering and non-linear interaction—to the neuropile between the axon terminals of lamina neurons L1 and L2 (ref. 28) and the dendrites of directionally selective T4 and T5 cells (Supplementary Fig. 2). The dendrites of T4 and T5 cells might well be the place where signals from neighbouring columns interact in a nonlinear way, similar to the dendrites of star-burst amacrine cells of the vertebrate retina[29].

METHODS SUMMARY

Flies. Flies used in calcium imaging experiments (Figs 1–3) had the following genotypes: T4/T5 line (w−; +/+; UAS-GCaMP5,R42F06-GAL4/UAS-GCaMP5, R42F06-GAL4), T4 line (w−; +/+; UAS-GCaMP5,R54A03-GAL4/UAS-GCaMP5, R54A03-GAL4), T5 line (w−; +/+; UAS-GCaMP5,R42H07-GAL4/UAS-GCaMP5, R42H07-GAL4). Flies used in electrophysiological and behavioural experiments (Fig. 4) had identical genotypes of the following kind: TNT control flies (w+/w+; UAS-TNT-E/UAS-TNT-E; +/+), T4 control flies (w+/w−; + /+; VT37588-GAL4/ +), T5 control flies (w+/w−; +/+; R42H07-GAL4/+), T4-block flies (w+/w−; UAS-TNT-E/+; VT37588-GAL4/+), T5-block flies (w+/w−; UAS-TNT-E/+; R42H07-GAL4/+).

Two-photon microscopy. We used a custom-built two-photon laser scanning microscope[29] equipped with a ×40 water immersion objective and a mode locked Ti:sapphire laser. To shield the photomultipliers from the stimulus light, two separate barriers were used: the first was placed directly over the LEDs, the second extended from the fly holder over the arena. Images were acquired at a resolution of 256 × 256 pixels and a frame rate of 1.87 Hz, except where indicated, using ScanImage software[30].

Electrophysiology. Recordings were established under visual control using a Zeiss Microscope and a ×40 water immersion objective.

Behavioural analysis. The locomotion recorder was custom-designed according to ref. 18. It consisted of an air-suspended sphere floating in a bowl-shaped sphere holder. Motion of the sphere was recorded by two optical tracking sensors.

Visual stimulation. For calcium imaging and electrophysiological experiments, we used a custom-built LED arena covering 180° and 90° of the visual field along the horizontal and the vertical axis, respectively, at 1.5° resolution. For the behavioural experiments, three 120-Hz LCD screens formed a U-shaped visual arena with the fly in the centre, covering 270° and 114° of the visual field along the horizontal and the vertical axes, respectively, at 0.1° resolution.

Data evaluation. Data were evaluated off-line using custom-written software (Matlab and IDL).

Full Methods and any associated references are available in the online version of the paper.

1. Bausenwein, B., Dittrich, A. P. M. & Fischbach, K. F. The optic lobe of *Drosophila melanogaster* II. Sorting of retinotopic pathways in the medulla. *Cell Tissue Res.* **267,** 17–28 (1992).
2. Akerboom, J. et al. Optimization of a GCaMP calcium indicator for neural activity imaging. *J. Neurosci.* **32,** 13819–13840 (2012).
3. Cajal, S. R. & Sanchez, D. *Contribucion al conocimiento de los centros nerviosos de los insectos* (Imprenta de Hijos de Nicholas Moja, 1915).
4. Strausfeld, N. J. *Atlas of an Insect Brain* (Springer, 1976).
5. Fischbach, K. F. & Dittrich, A. P. M. The optic lobe of *Drosophila melanogaster*. I. A Golgi analysis of wild-type structure. *Cell Tissue Res.* **258,** 441–475 (1989).

6. Reichardt, W. Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In *Sensory Communication* (ed. Rosenblith, W. A.) 303–317 (MIT Press and John Wiley & Sons, 1961).

7. Borst, A., Haag, J. & Reiff, D. F. Fly motion vision. *Annu. Rev. Neurosci.* **33,** 49–70 (2010).

8. Buchner, E., Buchner. S. & Buelthoff, l. Deoxyglucose mapping of nervous activity induced in *Drosophila* brain by visual movement. 1. Wildtype. *J. Comp. Physiol.* **155,** 471–483 (1984).

9. Strausfeld, N. J. & Lee, J. K. Neuronal basis for parallel visual processing in the fly. *Vis. Neurosci.* **7,** 13–33 (1991).

10. Schnell, B., Raghu, V. S., Nern, A. & Borst, A. Columnar cells necessary for motion responses of wide-field visual interneurons in *Drosophila. J. Comp. Physiol. A* **198,** 389–395 (2012).

11. Douglass, J. K. & Strausfeld, N. J. Visual motion-detection circuits in flies: Parallel direction- and non-direction-sensitive pathways between the medulla and lobula plate. *J. Neurosci.* **16,** 4551–4562 (1996).

12. Franceschini, N., Riehle, A. & Le Nestour, A. Directionally selective motion detection by insect neurons. In *Facets of Vision* (ed. Stavenga, H.) 360–390 (Springer, 1989).

13. Joesch, M., Schnell, B., Raghu, S. V., Reiff, D. F. & Borst, A. ON and OFF pathways in *Drosophila* motion vision. *Nature* **468,** 300–304 (2010).

14. Pfeiffer, B. D. *et al.* Tools for neuroanatomy and neurogenetics in *Drosophila. Proc. Natl Acad. Sci. USA* **105,** 9715–9720 (2008).

15. Denk, W., Strickler, J. H. & Webb, W. W. Two-photon laser scanning fluorescence microscopy. *Science* **248,** 73–76 (1990).

16. Oyster, C. W. & Barlow, H. B. Direction-selective units in rabbit retina: distribution of preferred directions. *Science* **155,** 841–842 (1967).

17. Sweeney, S. T., Broadie, K., Keane, J., Niemann, H. & O'Kane, C. J. Targeted expression of tetanus toxin light chain in *Drosophila* specifically eliminates synaptic transmission and causes behavioral defects. *Neuron* **14,** 341–351 (1995).

18. Seelig, J. D. *et al.* Two-photon calcium imaging from head-fixed *Drosophila* during optomotor walking behavior. *Nature Methods* **7,** 535–540 (2010).

19. Clark, D. A., Bursztyn, L., Horowitz, M. A., Schnitzer, M. J. & Clandinin, T. R. Defining the computational structure of the motion detector in *Drosophila. Neuron* **70,** 1165–1177 (2011).

20. Egelhaaf, M. & Borst, A. Calcium accumulation in visual interneurons of the fly: Stimulus dependence and relationship to membrane potential. *J. Neurophysiol.* **73,** 2540–2552 (1995).

21. Joesch, M., Plett, J., Borst, A. & Reiff, D. F. Response properties of motion-sensitive visual interneurons in the lobula plate of *Drosophila melanogaster. Curr. Biol.* **18,** 368–374 (2008).

22. Schnell, B. *et al.* Processing of horizontal optic flow in three visual interneurons of the *Drosophila* brain. *J. Neurophysiol.* **103,** 1646–1657 (2010).

23. Borst, A. & Egelhaaf, M. Direction selectivity of fly motion-sensitive neurons is computed in a two-stage process. *Proc. Natl Acad. Sci. USA* **87,** 9363–9367 (1990).

24. Single, S., Haag, J. & Borst, A. Dendritic computation of direction selectivity and gain control in visual interneurons. *J. Neurosci.* **17,** 6023–6030 (1997).

25. Eichner, H., Joesch, M., Schnell, B., Reiff, D. F. & Borst, A. Internal structure of the fly elementary motion detector. *Neuron* **70,** 1155–1164 (2011).

26. Joesch, M., Weber, F., Eichner, H. & Borst, A. Functional specialization of parallel motion detection circuits in the fly. *J. Neurosci.* **33,** 902–905 (2013).

27. Egelhaaf, M. & Borst, A. Are there separate ON and OFF channels in fly motion vision? *Vis. Neurosci.* **8,** 151–164 (1992).

28. Takemura, S. Y., Lu, Z. & Meinertzhagen, I. A. Synaptic circuits of the *Drosophila* optic lobe: the input terminals to the medulla. *J. Comp. Neurol.* **509,** 493–513 (2008).

29. Euler, T., Detwiler, P. B. & Denk, W. Directionally selective calcium signals in dendrites of starburst amacrine cells. *Nature* **418,** 845–852 (2002).

30. Pologruto, T. A., Sabatini, B. L. & Svoboda, K. ScanImage: Flexible software for operating laser scanning microscopes. *Biomed. Eng. Online* **2,** 13 (2003).

METHODS

Flies. Flies were raised on standard cornmeal-agar medium at 25 °C and 60% humidity throughout development on a 12 h light/12 h dark cycle. For calcium imaging, we used the genetically encoded single-wavelength indicator GCaMP5, variant G, with the following mutations: T302L, R303P and D380Y (ref. 2). Expression of GCaMP5 was directed by three different Gal4 lines, all from the Janelia Farm collection[14]. Flies used in calcium imaging experiments (Figs 1–3) had the following genotypes: T4/T5 line (w^-; +/+; UAS-GCaMP5,R42F06-GAL4/UAS-GCaMP5,R42F06-GAL4), T4 (w^-; +/+; UAS-GCaMP5,R54A03-GAL4/UAS-GCaMP5,R54A03-GAL4) and T5 line (w^-; +/+; UAS-GCaMP5,R42H07-GAL4/UAS-GCaMP5,R42H07-GAL4). All driver lines were generated by the methods described in ref. 14 and were identified by screening a database of imaged lines, followed by reimaging of selected lines[31]. As homozygous for both the Gal4-driver and the UAS-GCaMP5 genes, T4 flies also showed some residual expression in T5 cells, and T5 flies also in T4 cells. This unspecific expression, however, was in general less than 25% of the expression in the specific cells. Flies used in electrophysiological and behavioural experiments (Fig. 4) had identical genotypes of the following kind: TNT control flies (w^+/w^+; UAS-TNT-E/UAS-TNT-E; +/+), T4 control flies (w^+/w^-; + /+; VT37588-GAL4/+), T5 control flies (w^+/w^-; +/+; R42H07-GAL4/+), T4-block flies (w^+/w^-; UAS-TNT-E/+; VT37588-GAL4/+), T5-block flies (w^+/w^-; UAS-TNT-E/+; R42H07-GAL4/+). UAS-TNT-E flies were derived from the Bloomington Stock Center (stock no. 28837) and VT37588-Gal4 flies were derived from the VDRC (stock no. 205893). Before electrophysiological experiments, flies were anaesthetized on ice and waxed on a Plexiglas holder using bees wax. The dissection of the fly cuticle and exposure of the lobula plate were performed as described previously (for imaging experiments, see ref. 32; for electrophysiology, see ref. 21). Flies used in behavioural experiments were taken from 18 °C just before the experiment and immediately cold-anaesthetized. The head, the thorax and the wings were glued to a needle using near-ultraviolet bonding glue (Sinfony Opaque Dentin) and strong blue LED light (440 nm, dental curing-light, New Woodpecker).

Two-photon microscopy. We used a custom-built two-photon laser scanning microscope[33] equipped with a ×40 water immersion objective (0.80 NA, IR-Achroplan; Zeiss). Fluorescence was excited by a mode locked Ti:sapphire laser (<100 fs, 80 MHz, 700–1,020 nm; pumped by a 10 W CW laser; both Mai Tai; Spectraphysics) with a DeepSee accessory module attached for dispersion compensation control resulting in better pulse compression and fluorescence at the target sample. Laser power was adjusted to 10–20 mW at the sample, and an excitation wavelength of 910 nm was used. The photomultiplier tube (H10770PB-40, Hamamatsu) was equipped with a dichroic band-pass mirror (520/35, Brightline). Images were acquired at a resolution of 256 × 256 pixels and a frame rate of 1.87 Hz, except in Fig. 2 (7.5 Hz), using the ScanImage software[30].

Electrophysiology. Recordings were established under visual control using a ×40 water immersion objective (LumplanF, Olympus), a Zeiss microscope (Axiotech vario 100, Zeiss), and illumination (100 W fluorescence lamp, hot mirror, neutral density filter OD 0.3; all from Zeiss). To enhance tissue contrast, we used two polarization filters, one located as an excitation filter and the other as an emission filter, with slight deviation on their polarization plane. For eye protection, we additionally used a 420-nm LP filter on the light path.

Behavioural analysis. The locomotion recorder was custom-designed according to ref. 18. Briefly, it consists of an air-suspended sphere floating in a bowl-shaped sphere holder. A high-power infrared LED (800 nm, JET series, 90 mW, Roithner Electronics) is located in the back to illuminate the fly and the sphere surface. Two optical tracking sensors are equipped with lens and aperture systems to focus on the sphere behind the fly. The tracking data are processed at 4 kHz internally, read out via a USB interface and processed by a computer at ≈200 Hz. This allows real-time calculation of the instantaneous rotation axis of the sphere. A third camera (GRAS-20S4M-C, Point Grey Research) is located in the back which is essential for proper positioning of the fly and allows real-time observation and video recording of the fly during experiments.

Visual stimulation. For calcium imaging and electrophysiological experiments, we used a custom-built LED arena that allowed refresh rates of up to 550 Hz and 16 intensity levels. It covered 180° (1.5° resolution) and 90° (1.5° resolution) of the visual field along the horizontal and the vertical axis, respectively. The LED arena was engineered and modified based upon ref. 34. The LED array consists of 7 × 4 individual TA08-81GWA dot-matrix displays (Kingbright), each harbouring 8 × 8 individual green (568 nm) LEDs. Each dot-matrix display is controlled by an ATmega168 microcontroller (Atmel) combined with a ULN2804 line driver (Toshiba America) acting as a current sink. All panels are in turn controlled via an I2C interface by an ATmega128 (Atmel)-based main controller board, which reads in pattern information from a compact flash (CF) memory card. Matlab was used for programming and generation of the patterns as well as for sending the serial command sequences via RS-232 to the main controller board. The luminance range of the stimuli was 0.5–33 cd m^{-2}. For the calcium imaging experiments, two separate barriers were used to shield the photomultipliers from the stimulus light coming from the LED arena. The first was a spectral filter with transparency to wavelengths >540 nm placed directly over the LEDs (ASF SFG 10, Microchemicals). The second was a layer of black PVC extending from the fly holder over the arena. Square wave gratings had a spatial wavelength of 30° of visual angle and a contrast of 88%. Unless otherwise stated, they were moving at 30° s^{-1}. Edges had the same contrast and were also moving at 30° s^{-1}. For the experiments shown in Figs 1, 2b and 3, each grating or edge motion was shown twice within a single sweep, resulting in a total of eight stimulation periods. Each stimulus period lasted 4 s, and subsequent stimuli were preceded by a 3-s pause. In the experiment shown in Fig. 2a, a dark edge of 88% contrast was moved for 1 s at 15° s^{-1} from the front to the back at three different positions (22°, 44°, 66°, from frontal to lateral). At each position, edge motion was repeated 15 times. For the experiment shown in Fig. 2b, a bright edge of 88% contrast was moving at 15° s^{-1} from the back to the front, and images were acquired at a frame rate of 7.5 Hz. For the experiments shown in Figs 3e, f, all six stimulus velocities were presented once within one sweep, with the stimulus lasting 4 s, and different stimuli being separated by 2 s. In the experiments shown in Figs 3g, h, a single sweep contained all 12 grating orientations with the same stimulus and pause length as above. For the electrophysiology experiments (Fig. 4a–f), multiple edges were used as stimuli moving simultaneously at 50° s^{-1}. To stimulate cells of horizontal system (HS cells), a vertical, stationary square-wave grating with 45° spatial wavelength was presented. For ON-edge motion, the right (preferred direction, PD) or the left edge (null direction, ND) of each light bar started moving until it merged with the neighbouring bar. For OFF-edge motion, the right or the left edge of each dark bar was moving. To stimulate cells of the vertical system (VS cells), the pattern was rotated by 90° clockwise. For the behavioural experiments (Fig. 4g–i), three 120-Hz LCD screens (Samsung 2233 RZ) were vertically arranged to form a U-shaped visual arena (w = 31 cm × d = 31 cm × h = 47 cm) with the fly in the centre. The luminance ranged from 0 to 131 cd m^{-2} and covered large parts of the flies' visual field (horizontal, ±135°; vertical, ±57°; resolution, <0.1°). The three LCD screens were controlled via NVIDIA 3D Vision Surround Technology on Windows 7 64-bit allowing a synchronized update of the screens at 120 frames per second. Visual stimuli were created using Panda3D, an open-source gaming engine, and Python 2.7, which simultaneously controlled the frame rendering in Panda3D, read out the tracking data and temperature and streamed data to the hard disk. The balanced motion stimulus consisted of a square-wave grating with 45° spatial wavelength and a contrast of 63%. Upon stimulation onset, dark and bright edges moved into opposite directions at 10° s^{-1} for 2.25 s. This stimulation was performed for both possible edge directions and two initial grating positions shifted by half a wavelength, yielding a total of four stimulus conditions.

Data evaluation. Data were evaluated off-line using custom-written software (Matlab and IDL). For the images shown in Figs 1e, f, 2a and 3a, b, the raw image series was converted into four images representing the relative fluorescence change during each direction of grating motion: $(\Delta F/F)_{stim} = (F_{stim} - F_{ref})/F_{ref}$. The image representing the stimulus fluorescence (F_{stim}) was obtained by averaging all images during stimulation; the image representing the reference fluorescence (F_{ref}) was obtained by averaging three images before stimulation. Both images were smoothed using a Gaussian filter of 10 pixel half-width. For the images shown in Figs 1f and 3a, b, $\Delta F/F$ images were normalized by their maximum value. Then, a particular colour was assigned to each pixel according to the stimulus direction during which it reached maximum value, provided it passed a threshold of 25%. Otherwise, it was assigned to background. The response strength of each pixel was coded as the saturation of that particular colour. For the data shown in Figs 2b, c and 3c–h, the raw image series was first converted into a $\Delta F/F$ series by using the first three images as reference. Then, a region was defined within a raw image, and average $\Delta F/F$ values were determined within that region for each image, resulting in a $\Delta F/F$ signal over time. Responses were defined as the maximum $\Delta F/F$ value reached during each stimulus presentation minus the average $\Delta F/F$ value during the two images preceding the stimulus. For the bar graphs shown in Fig. 4c, f, the average voltage responses during edge motion (0.45 s) along the cell's preferred (PD) and null direction (ND) were calculated. For each recorded tangential cell, the difference between the PD and the ND response was determined, and these values were averaged across all recorded cells. The data shown in Fig. 4g, h were obtained from the four stimulus conditions by averaging the turning responses for the two starting positions of the grating and calculating the mean difference between the turning responses for the two edge directions. For the bar graph shown in Fig. 4i, the average turning response of each fly during the last second of balanced motion stimulation was calculated. These values were averaged across all recorded flies within each genotype.

31. Jenett, A. *et al.* A Gal4-driver line resource for *Drosophila* neurobiology. *Cell Rep.* **2,** 991–1001 (2012).
32. Reiff, D. F., Plett, J., Mank, M., Griesbeck, O. & Borst, A. Visualizing retinotopic half-wave rectified input to the motion detection circuitry of *Drosophila. Nature Neurosci.* **13,** 973–978 (2010).
33. Euler, T. *et al.* Eyecup scope—optical recording of light stimulus-evoked fluorescence signals in the retina. *Pfluger Arch.* **457,** 1393–1414 (2009).
34. Reiser, M. B. & Dickinson, M. H. A modular display system for insect behavioral neuroscience. *J. Neurosci. Methods* **167,** 127–139 (2008).

LETTER

Generation of inner ear sensory epithelia from pluripotent stem cells in 3D culture

Karl R. Koehler[1,2,3], Andrew M. Mikosz[2,3], Andrei I. Molosh[2,4], Dharmeshkumar Patel[2,3] & Eri Hashino[1,2,3]

The inner ear contains sensory epithelia that detect head movements, gravity and sound. It is unclear how to develop these sensory epithelia from pluripotent stem cells, a process that will be critical for modelling inner ear disorders or developing cell-based therapies for profound hearing loss and balance disorders[1,2]. So far, attempts to derive inner ear mechanosensitive hair cells and sensory neurons have resulted in inefficient or incomplete phenotypic conversion of stem cells into inner-ear-like cells[3–7]. A key insight lacking from these previous studies is the importance of the non-neural and preplacodal ectoderm, two critical precursors during inner ear development[8–11]. Here we report the stepwise differentiation of inner ear sensory epithelia from mouse embryonic stem cells (ESCs) in three-dimensional culture[12,13]. We show that by recapitulating *in vivo* development with precise temporal control of signalling pathways, ESC aggregates transform sequentially into non-neural, preplacodal and otic-placode-like epithelia. Notably, in a self-organized process that mimics normal development, vesicles containing prosensory cells emerge from the presumptive otic placodes and give rise to hair cells bearing stereocilia bundles and a kinocilium. Moreover, these stem-cell-derived hair cells exhibit functional properties of native mechanosensitive hair cells and form specialized synapses with sensory neurons that have also arisen from ESCs in the culture. Finally, we demonstrate how these vesicles are structurally and biochemically comparable to developing vestibular end organs. Our data thus establish a new *in vitro* model of inner ear differentiation that can be used to gain deeper insight into inner ear development and disorder.

During neurulation *in vivo*, the definitive ectoderm is subdivided into the neuroectoderm and non-neural ectoderm, the latter of which gives rise to the inner ear (Supplementary Fig. 1a). Recent studies have demonstrated how organogenesis of complex neuroectoderm tissues such as the cerebral cortex and retina can be faithfully reconstituted *in vitro* by culturing ESCs as a floating aggregate in serum-free media (serum-free floating culture of embryoid body-like aggregates with quick reaggregation; SFEBq culture)[12,14,15]. As the inner ear shares a common precursor with these tissues, the definitive ectoderm, we proposed that SFEBq culture could be redirected to generate inner ear epithelia using carefully timed morphogenetic cues (Fig. 1a and Supplementary Fig. 1b). Led by previous studies, we identified a definitive-ectoderm-like epithelium on day 3 of SFEBq culture, before the expression of neuroectoderm-associated proteins on day 5 (Supplementary Fig. 1c–j)[13,16]. During early embryogenesis, activation of bone morphogenetic protein (BMP) signalling is critical for induction of the non-neural ectoderm from the definitive ectoderm epithelium[11,17]. Consistent with this role, in aggregates treated with human BMP4 (hereafter BMP), the non-neural ectoderm marker *Dlx3* was upregulated, whereas the neuroectoderm marker *Sox1* was downregulated (Supplementary Fig. 1k, l). However, BMP-treated aggregates also expressed the mesendoderm marker brachyury (also known as *T*), indicating the undesirable induction of mesoderm or endoderm cell types (Fig. 1b–h and Supplementary Fig. 2b)[18]. To

suppress aberrant mesendoderm induction, we combined BMP treatment with the transforming growth factor-β (TGF-β) inhibitor SB-431542 (SB; Fig. 1a)[19]. A combined treatment of SB and BMP (BMP/SB) on day 3 completely abolished brachyury[+] cells in the outer epithelium (Fig. 1d, e, h; see also Supplementary Discussion)[12,18,19].

To test whether BMP/SB treatment indeed induced non-neural ectoderm, we assessed the cellular composition of BMP/SB-treated aggregates by immunofluorescence at differentiation day 5. Notably, expression of the non-neural ectoderm marker activator protein 2 (AP2, also known as Tfap2a) was found predominantly in the E-cadherin (Ecad, also known as Cdh1)[+] outer epithelium, but was absent in other regions of treated aggregates (Fig. 1i, j). Moreover, we identified an intermediate layer of
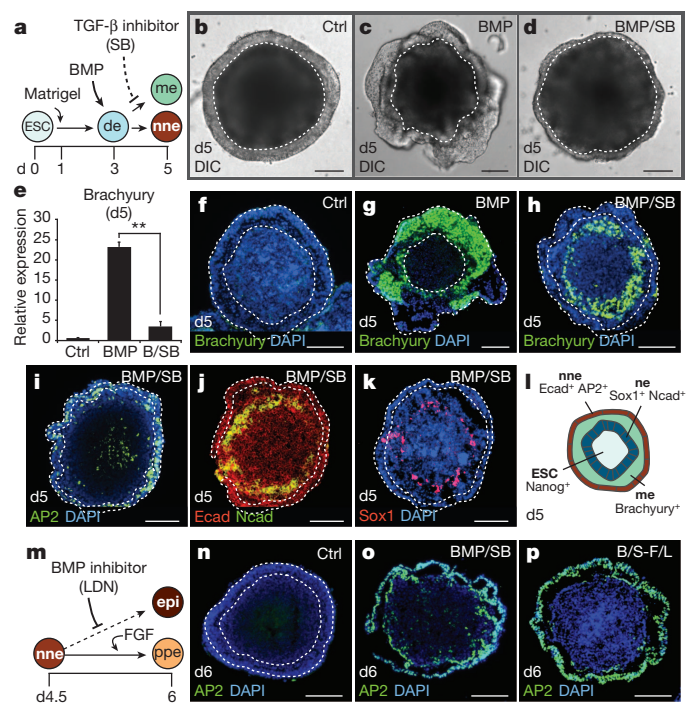

Figure 1 | Non-neural and preplacodal ectoderm induction in three-dimensional culture. a, Non-neural ectoderm induction strategy. d, day; de, definitive ectoderm; me, mesendoderm; nne, non-neural ectoderm. **b–d**, Morphology of control (Ctrl), BMP and BMP/SB aggregates. DIC, differential interference contrast. **e**, SB decreases the level of brachyury expression induced by BMP ($n = 3$; **$P < 0.01$; mean ± s.e.m.). **f–h**, Brachyury[+] cells are less prevalent in BMP/SB aggregates. DAPI, 4′,6-diamidino-2-phenylindole. **i–k**, BMP/SB aggregates contain an outer AP2[+] Ecad[+] epithelium and an interior Sox1[+] Ncad[+] cell layer. **l**, BMP/SB aggregate composition on day 5. ne, neuroectoderm. **m**, Preplacodal ectoderm induction strategy. epi, epidermis; ppe, preplacodal ectoderm. **n–p**, BMP/SB-FGF/LDN (B/S-F/L) aggregates are distinguished by a thickened AP2[+] epithelium absent in other conditions. Scale bars, 100 μm.

[1]Medical Neuroscience Graduate Program, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA. [2]Stark Neurosciences Research Institute, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA. [3]Department of Otolaryngology-Head and Neck Surgery, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA. [4]Department of Psychiatry, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA.

each aggregate with Sox1[+] and N-cadherin (Ncad, also known as Cdh2)[+] cells, indicative of the formation of neuroectoderm (Fig. 1j, k and Supplementary Fig. 1h). In addition, the pluripotency marker Nanog was confined to cells at the core of each aggregate (Supplementary Fig. 2g). Altogether, these data strongly suggest that the outer epithelium of day 5 BMP/SB-treated aggregates represents nonneural ectoderm, which surrounds an interior layer containing a mixture of mesendodermal and neuroectodermal tissues and a central core of pluripotent cells (Fig. 1l and Supplementary Fig. 2). In support of this conclusion, the outer epithelium of BMP/SB samples develops into a Krt5[+] p63 (also known as Trp63)[+] epithelium, mimicking embryonic development of the epidermis (Supplementary Fig. 3).

The preplacodal region, a contiguous band of embryonic head ectoderm, arises from the non-neural ectoderm at the neural tube border and is the precursor to all of the cranial placodes (Supplementary Fig. 4a)[11]. Although BMP signalling is required for induction of non-neural ectoderm, recent studies suggest that subsequent BMP inhibition, along with active fibroblast growth factor (FGF) signalling, is necessary for non-neural cells to select a preplacodal over an epidermal fate (Fig. 1m)[9,20,21]. With this in mind, we began treating BMP/SB aggregates with various combinations of the specific BMP inhibitor LDN-193189 (LDN) and human FGF2. We found that BMP/SB aggregates treated with LDN on day 4.5 maintained expression of *Dlx3*, indicating that BMP inhibition after non-neural induction does not reverse cell fate specification (Supplementary Fig. 4c). Consistent with the preplacodal ectoderm being thickened relative to the surrounding surface

ectoderm[8,11], we found thickened patches of epithelia in BMP/SB/LDN aggregates that were not present in BMP/SB epithelia (Supplementary Fig. 4b, h). As observed *in vivo*, this morphological change appeared to be dependent on endogenous FGFs, as inhibition of FGF signalling by the small molecule SU5402 abolished epithelial thickening (Supplementary Fig. 5a). A combined treatment of recombinant FGF2 and LDN (hereafter BMP/SB-FGF/LDN) significantly increased the thickness of the epithelium compared to BMP/SB and BMP/SB-LDN aggregates (Fig. 1n–p and Supplementary Fig. 4h). Notably, in >95% of BMP/SB-FGF/LDN aggregates, a thickened Gata3[+] Six1[+] AP2[+] epithelium ruffled and formed ovoid vesicles between days 6 and 8 (Fig. 1p and Supplementary Fig. 4d–j). These and the following data strongly suggest that the outer epithelium of BMP/SB-FGF/LDN-treated aggregates is representative of preplacodal ectoderm.

In vertebrates, the otic placode is derived from a posterior preplacodal region known as the otic-epibranchial placode domain (OEPD; Fig. 2a). The otic placode is demarcated from other developing placodes by expression of the transcription factors Pax2 and Pax8 (see Supplementary Fig. 6 for the *in vivo* situation)[22]. Because the induction of the OEPD requires FGF signalling and the otic placode epithelium thickens, invaginates and forms the otic vesicle[10], we examined whether the vesicle-forming epithelia of BMP/SB-FGF/LDN aggregates were representative of the primordial inner ear. Our quantitative PCR analysis revealed that *Pax2* and *Pax8* were significantly upregulated in BMP/SB-FGF/LDN samples compared to other conditions (Fig. 2b, c). By day 6, we observed Pax8[+] cells distributed in placode-like patches throughout the outer Ecad[+] epithelium of BMP/SB-FGF/LDN aggregates only (Fig. 2d, e). Notably, we also observed a population of Pax8[+] Ecad[−] cells in the interior of each aggregate, suggesting formation of mid-hindbrain tissue in this region (Supplementary Fig. 7). The percentage of Pax8[+] Ecad[+] epithelium markedly increased between days 6 and 8 (Fig. 2f, g and Supplementary Fig. 8a–e) and the Pax8[+] Ecad[+] epithelium bore a striking morphological resemblance to the developing otic placode (Supplementary Fig. 6). Of note, we did not observe expression of Pax3 or Pax6 in the outer epithelium, ruling out the development of other cranial placodes (Supplementary Fig. 7c–g). Taken together, these findings show that FGF/LDN treatment is critically important for *in vitro* otic placode induction and that treatment is most effective when performed between days 4 and 5 (Supplementary Fig. 8).

In vivo, the prosensory domain of the otic placode/vesicle gives rise to the vestibular/cochlear sensory epithelia and inner ear sensory neurons. Otic prosensory cells are defined by the expression of Pax2, Pax8, Ecad, Sox2 and jagged 1 (Jag1) (Supplementary Figs 6 and 12). On day 8 of differentiation, BMP/SB-FGF/LDN aggregates were transferred to a serum-free floating culture to allow self-guided differentiation. In each aggregate analysed, approximately 24 h after transfer, the interior cell mass breached the outer epithelium and formed a heterogeneous cell layer on the exterior of the aggregate (n = 253 aggregates; Fig. 2h, m and Supplementary Fig. 9). This indicated that the outer epithelium transitions to an inner epithelium lining the core of each aggregate. During days 9–12 we observed the continuous evagination of vesicles containing Pax2[+] Ecad[+], Pax2[+] Pax8[+], and Sox2[+] Pax8[+] cells from the presumptive OEPD epithelium into the exterior cell layer (Fig. 2h–k and Supplementary Figs 9 and 10), which resulted in ~25 Pax2[+] Pax8[+] Sox2[+] vesicles per aggregate (Fig. 2l). We proposed that endogenous Wnt signalling may underlie induction of vesicles bearing otic prosensory markers in our culture because Wnt signalling is necessary for otic placode formation *in vivo*[10]. Confirming this hypothesis, treatment of aggregates with the Wnt inhibitor XAV939 from days 8–10 significantly decreased the number of prosensory vesicles and, specifically, reduced the prevalence of Pax2[+] vesicles (Fig. 2l). These data indicate that endogenous Wnt signalling induces the formation of otic vesicles from the presumptive otic placode using similar mechanisms as observed *in vivo*. Interestingly, the remaining inner epithelium developed into Krt5[+] p63[+] epidermis and the exterior layer of cells gave rise

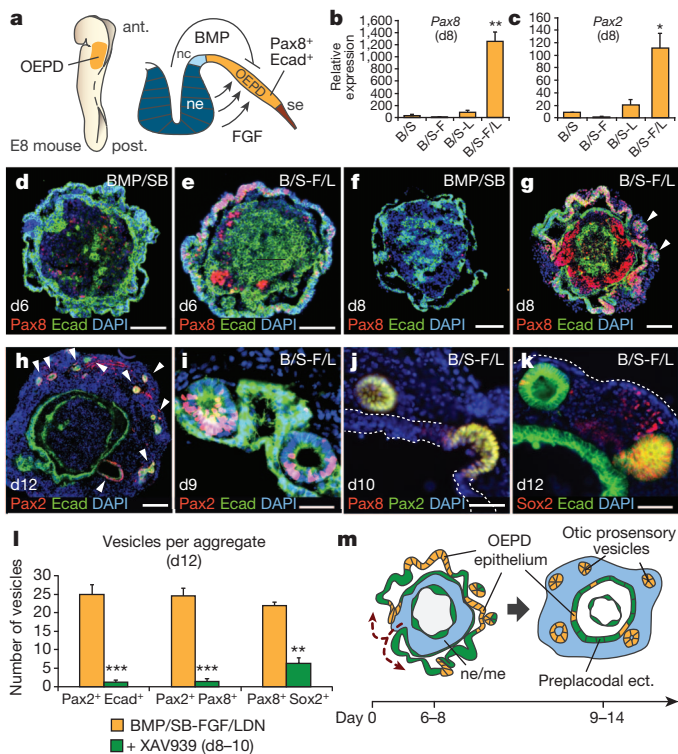**Figure 2 | Otic induction from the preplacodal epithelium *in vitro*. a**, OEPD induction in mice. nc, neural crest; se, surface ectoderm. **b, c**, *Pax8* (**b**) *Pax2* (**c**) messenger RNA expression on day 8 (n = 3–4; *P < 0.05, **P < 0.01; mean ± s.e.m.). **d–g**, Pax8 and Ecad expression in BMP/SB aggregates (**d, f**) and BMP/SB-FGF/LDN aggregates (**e, g**) on days 6 and 8. Arrowheads indicate vesicles. **h**, Day 12 BMP/SB-FGF/LDN aggregate with Pax2[+] Ecad[+] vesicles (arrowheads). **i–k**, Pax2[+] Ecad[+] (**i**), Pax2[+] Pax8[+] (**j**) and Sox2[+] Ecad[+] (**k**) vesicles invaginate from the inner epithelium from days 9–12. **l**, XAV939 decreases the number of vesicles expressing Pax2 and Ecad, Pax2 and Pax8, and Pax8 and Sox2 on day 12. (n = 9 aggregates; **P < 0.01, ***P < 0.001; mean ± s.e.m.). **m**, Self-guided, inside-out rearrangement of BMP/SB-FGF/LDN aggregates and formation of otic vesicles. Scale bars, 100 μm (**d–f, h**), 50 μm (**i–k**) and 25 μm (**g**).

to mesenchyme-tissue-like cartilage and adipose (Supplementary Fig. 11). The basal (p63$^+$) layer of the inner epithelium was oriented so that the apical surface of the epithelium was facing the interior of the aggregate. Thus, the process of vesicle evagination towards the outside of the aggregate is consistent with the orientation of embryonic otic vesicle invagination into the head mesenchyme (Supplementary Fig. 11).

During development, the prosensory domain of the otic vesicle is destined to become sensory epithelia harbouring Myo7a$^+$ sensory hair cells. Surprisingly, the epithelia of Sox2$^+$ Jag1$^+$ vesicles became Myo7a$^+$ by day 14, mimicking the diffuse Myo7a staining pattern in the embryonic day (E)9.5 otic vesicle (Fig. 3a, b and Supplementary Fig. 12a–f). By day 16, we found that each aggregate contained 15.4 ± 4.8 ($n = 12$ aggregates) vesicles lined with Myo7a$^+$ Sox2$^+$ cells bearing the stereotyped morphology of sensory hair cells with a large nuclei (~8-μm diameter) positioned basal to an elongated apical end (Fig. 3c). The Myo7a$^+$ Sox2$^+$ cells were organized in a radial pattern with the apical end abutting a lumen of varying sizes (~5–1,100-μm-long-axis diameter; Fig. 3c–e). Basal to each layer of Myo7a$^+$ Sox2$^+$ cells was a tightly arranged layer of Sox2$^+$ cells reminiscent of supporting cells (Fig. 3c–j and Supplementary Video 1). Mimicking the in vivo sensory epithelia, hair cells and supporting cells could be further distinguished by expression of Brn3c (also known as Pou4f3) and cyclin D1, respectively (Supplementary Fig. 13a–f)[23]. F-actin staining revealed tight cell–cell junctions along the luminal surface as well as F-actin$^+$ espin (Espn)$^+$ stereocilia bundles (Fig. 3k–o, Supplementary Fig. 13g–i and Supplementary Video 2). Every Myo7a$^+$ cell analysed also had an acetylated-α-tubulin$^+$ kinocilium protruding from the

apical end into the lumen (Fig. 3m, n and Supplementary Fig. 13j–m). Stereocilia and kinocilium were not visible at day 16, but the average height increased from day 20 to day 24 and fell within the range of heights recorded from an adult mouse utricle (Fig. 3o)[24]. The hair cells also appeared to be functional on the basis of rapid uptake of FM1-43 dye and the diversity of voltage-dependent currents (Fig. 3p–r and Supplementary Fig. 14)[25,26]. In all cells included in this study we observed outwardly rectifying potassium currents with voltage-dependent activation kinetics to amplitudes ranging from 194 pA to 3,612 pA with a mean of 1,003 ± 527 pA ($n = 6$; Fig. 3r). In addition, some cells were distinguished by the presence of a transient inward current, probably reflecting calcium channel activity (Supplementary Fig. 14k, l). By day 20 each BMP/SB-FGF/LDN aggregate contained 1,552.3 ± 83.1 Myo7a$^+$ cells with typical hair cell morphology, in marked contrast to other conditions that yielded no Myo7a$^+$ cells (~1–2% of all cells in the aggregate; $n = 12$–16 aggregates per condition; Fig. 3s, t). We conclude from these data that the cytoarchitecture, cellular morphology and functional characteristics observed in Myo7a$^+$ Sox2$^+$ vesicles are identical to sensory epithelia in the inner ear.

There are four distinct populations of hair cells in the mammalian inner ear; type I and type II vestibular and inner and outer cochlear hair cells. We wished to reveal which type of hair cells populated the stem-cell-derived sensory epithelia in our culture. Previous studies have suggested that expression of Pax2 and Sox2 may distinguish vestibular from cochlear hair cells[27,28]. In addition, expression of the calcium-binding protein calbindin 2 (Calb2) and Sox2 uniquely label type II vestibular hair cells, whereas calyceal innervation from sensory
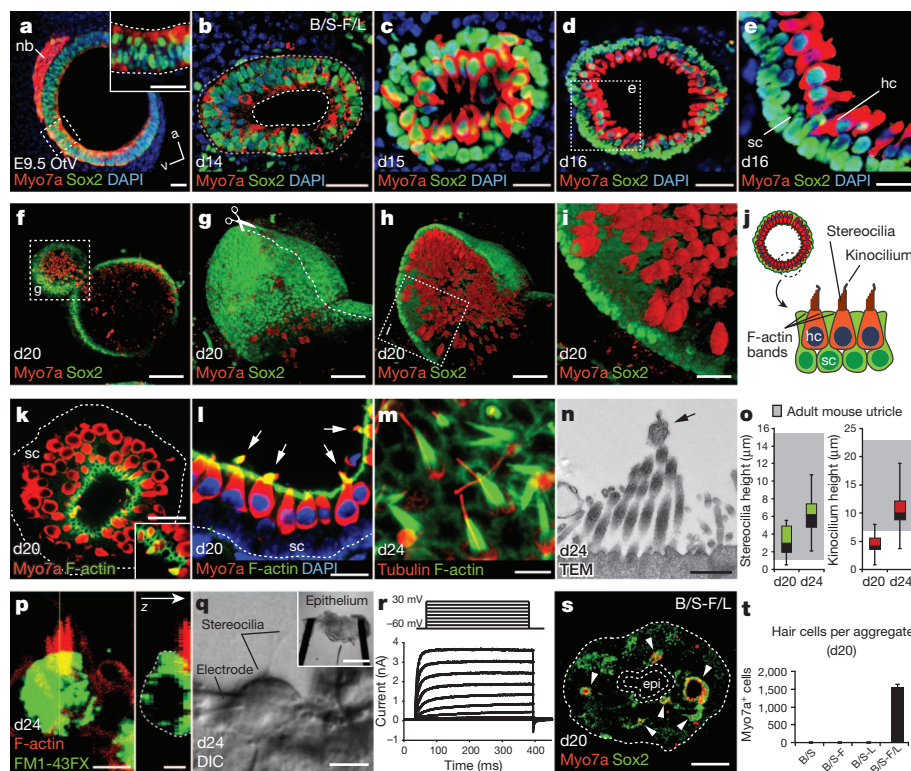

Figure 3 | Stem-cell-derived otic vesicles generate functional inner ear hair cells. a, b, Expression of Myo7a in the E9.5 otic vesicle (OtV) (**a**) and day 14 vesicles (**b**). nb, neuroblasts. **c–e,** Myo7a$^+$ Sox2$^+$ hair cells (hc) with underlying Sox2$^+$ supporting cells (sc) on days 15 (**c**) and 16 (**d, e**). **f–i,** Whole-mount immunofluorescence for Myo7a and Sox2 (**f**) and three-dimensional reconstruction (**g–i**) of a vesicle in a day 20 BMP/SB-FGF/LDN aggregate. **j,** Vesicles display the hallmarks of inner ear sensory epithelia. **k–m,** F-actin labels cell–cell junctions on the luminal surface and stereocilia bundles. **m,** Acetylated α-tubulin (tubulin) labels kinocilium and the cuticular plate. **n,** Transmission electron micrograph (TEM) of stereocilia bundles and kinocilium (arrow). **o,** Distribution of stereocilia and kinocilium heights on

days 20 and 24 compared to adult mouse utricle, with range indicated by grey boxes ($n > 100$ cells; ± max/min). **p,** Representative hair cell following 1-min FM1-43FX incubation, fixation and staining for F-actin. **q,** Representative epithelium preparation (inset) and hair cell during electrophysiological recordings. **r,** Representative voltage–current responses recorded from hair cells. The voltage protocol is shown at the top. **s,** Day 20 aggregate with Myo7a$^+$ Sox2$^+$ vesicles. Epidermis is indicated by dashed outline. **t,** Number of hair cells on day 20 ($n = 12$–16; mean ± s.e.m.). Scale bars, 250 μm (**f, s, q** (inset)), 50 μm (**d, g, h**), 25 μm (**a–c, e, i, k, l**), 10 μm (**q**), 5 μm (**m, p**) and 250 nm (**n**).

neurons identifies type I vestibular hair cells (Fig. 4a)[27,29,30]. On day 20, nearly all stem-cell-derived hair cells were Sox2$^+$ Pax2$^+$ ($n > 250$ hair cells; Figs 3c–e, i, 4b). Moreover, every hair cell expressed Calb2, suggesting a uniform population of type II vestibular hair cells (Fig. 4c, e). From a structural standpoint, we noted the presence of larger lumen vesicles (3.7 ± 0.3 per aggregate, defined as >50-μm-long-axis diameter, $n = 15$ aggregates) with regions of sensory (with hair cells) and nonsensory (without hair cells) epithelia identical in organization to a vestibular end organ (Fig. 4d–g and Supplementary Figs 14b, h and 15). Intriguingly, we also observed discrete populations of Calb2$^+$ and Brn3a$^+$ Tuj1 (also known as Tubb3)$^+$ neurofilament (Nefl)$^+$ neurons extended processes towards the sensory epithelia (Fig. 4h and Supplementary Fig. 16a–d). We were surprised to find that, by day 16, hair cells exhibited punctate expression of ribeye (Ctbp2) colocalized with the postsynaptic and neuronal markers Tuj1, synaptophysin (Syp), Snap25 and Rab3a, indicating the formation of ribbon synapses with adjacent neurons (Fig. 4h–j and Supplementary Fig. 16e–j). Notably, the number of ribbon synapses increased over time in culture, suggesting a maturation process similar to native hair cells in the inner ear (Fig. 4k and Supplementary Fig. 16h–j)[30]. Together, these results indicate that stem-cell-derived vesicles in our culture represent immature vestibular end organs, specifically the utricle and/or saccule[12,23,28].

In conclusion, our findings highlight a binary mechanism of BMP activation and TGF-β inhibition underlying *in vitro* non-neural ectoderm induction. Furthermore, subsequent inhibition of BMP signalling concomitant with activation of FGF signalling are required for preplacodal induction. Notably, the formation of these precursors is sufficient to trigger self-guided induction of the sensory epithelia, from which hair cells with structural and functional properties of native mechanosensitive hair cells in the inner ear spontaneously arise in significant numbers (~1,500 hair cells per aggregate; see Fig. 4l). This new approach can not only be used as a potent model system to elucidate the mechanisms underlying inner ear development, but will also provide an easily accessible and reproducible means of generating hair cells for *in vitro* disease modelling, drug discovery or cellular therapy experiments.

METHODS SUMMARY

The SFEBq culture system was performed as described previously, but with major modifications[12]. On day 3 of the protocol, BMP4 ($10 \, \mathrm{ng \, ml^{-1}}$) and SB-431542 ($1 \, \mu M$) were added to each well at 5× concentration in 25 μl of fresh media. On day 4.5, FGF2 ($25 \, \mathrm{ng \, ml^{-1}}$) and LDN-193189 ($100 \, \mathrm{nM}$) were added to each well at 6× concentration in 25 μl of fresh media.

Full Methods and any associated references are available in the online version of the paper.

1. Bermingham-McDonogh, O. & Reh, T. A. Regulated reprogramming in the regeneration of sensory receptor cells. *Neuron* **71**, 389–405 (2011).
2. Brigande, J. V. & Heller, S. *Quo vadis*, hair cell regeneration? *Nature Neurosci.* **12**, 679–685 (2009).
3. Ouji, Y., Ishizaka, S., Nakamura-Uchiyama, F. & Yoshikawa, M. *In vitro* differentiation of mouse embryonic stem cells into inner ear hair cell-like cells using stromal cell conditioned medium. *Cell Death Dis.* **3**, e314 (2012).
4. Oshima, K. *et al.* Mechanosensitive hair cell-like cells from embryonic and induced pluripotent stem cells. *Cell* **141**, 704–716 (2010).
5. Kondo, T. *et al.* Tlx3 exerts context-dependent transcriptional regulation and promotes neuronal differentiation from embryonic stem cells. *Proc. Natl Acad. Sci. USA* **105**, 5780–5785 (2008).
6. Reyes, J. H. *et al.* Glutamatergic neuronal differentiation of mouse embryonic stem cells after transient expression of neurogenin 1 and treatment with BDNF and GDNF: *in vitro* and *in vivo* studies. *J. Neurosci.* **28**, 12622–12631 (2008).
7. Chen, W. *et al.* Restoration of auditory evoked responses by human ES-cell-derived otic progenitors. *Nature* **490**, 278–282 (2012).
8. Schlosser, G. Induction and specification of cranial placodes. *Dev. Biol.* **294**, 303–351 (2006).
9. Pieper, M., Ahrens, K., Rink, E., Peter, A. & Schlosser, G. Differential distribution of competence for panplacodal and neural crest induction to non-neural and neural ectoderm. *Development* **139**, 1175–1187 (2012).
10. Groves, A. K. & Fekete, D. M. Shaping sound in space: the regulation of inner ear patterning. *Development* **139**, 245–257 (2012).
11. Grocott, T., Tambalo, M. & Streit, A. The peripheral sensory nervous system in the vertebrate head: a gene regulatory perspective. *Dev. Biol.* **370**, 3–23 (2012).
12. Eiraku, M. *et al.* Self-organizing optic-cup morphogenesis in three-dimensional culture. *Nature* **472**, 51–56 (2011).
13. Suga, H. *et al.* Self-formation of functional adenohypophysis in three-dimensional culture. *Nature* **480**, 57–62 (2011).
14. Eiraku, M. *et al.* Self-organized formation of polarized cortical tissues from ESCs and its active manipulation by extrinsic signals. *Cell Stem Cell* **3**, 519–532 (2008).
15. Nakano, T. *et al.* Self-formation of optic cups and storable stratified neural retina from human ESCs. *Cell Stem Cell* **10**, 771–785 (2012).
16. Kamiya, D. *et al.* Intrinsic transition of embryonic stem-cell differentiation into neural progenitors. *Nature* **470**, 503–509 (2011).
17. Wilson, P. A. & Hemmati-Brivanlou, A. Induction of epidermis and inhibition of neural fate by Bmp-4. *Nature* **376**, 331–333 (1995).
18. Bernardo, A. S. *et al.* BRACHYURY and CDX2 mediate BMP-induced differentiation of human and mouse pluripotent stem cells into embryonic and extraembryonic lineages. *Cell Stem Cell* **9**, 144–155 (2011).
19. Chambers, S. M. *et al.* Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nature Biotechnol.* **27**, 275–280 (2009).
20. Kwon, H.-J., Bhat, N., Sweet, E. M., Cornell, R. A. & Riley, B. B. Identification of early requirements for preplacodal ectoderm and sensory organ development. *PLoS Genet.* **6**, e1001133 (2010).
21. Kwon, H.-J. & Riley, B. B. Mesendodermal signals required for otic induction: Bmp-antagonists cooperate with Fgf and can facilitate formation of ectopic otic tissue. *Dev. Dyn.* **238**, 1582–1594 (2009).
22. Ladher, R. K., O'Neill, P. & Begbie, J. From shared lineage to distinct functions: the development of the inner ear and epibranchial placodes. *Development* **137**, 1777–1785 (2010).
23. Laine, H., Sulg, M., Kirjavainen, A. & Pirvola, U. Cell cycle regulation in the inner ear sensory epithelia: role of cyclin D1 and cyclin-dependent kinase inhibitors. *Dev. Biol.* **337**, 134–146 (2010).
24. Li, A., Xue, J. & Peterson, E. H. Architecture of the mouse utricle: macular organization and hair bundle heights. *J. Neurophysiol.* **99**, 718–733 (2008).
25. Meyers, J. R. *et al.* Lighting up the senses: FM1–43 loading of sensory cells through nonselective ion channels. *J. Neurosci.* **23**, 4054–4065 (2003).

Figure 4 | Stem-cell-derived sensory epithelia are comparable to immature vestibular end organs. a, Schematic of vestibular end organs and type I/II vestibular hair cells. vgn, vestibular ganglion neurons. **b, c**, Pax2 (**b**) and Calb2 (**c**) are expressed in all Myo7a$^+$ stem-cell-derived hair cells on day 20. CyclinD1 (cD1) is expressed in supporting cells. **d–g**, The structural organization of vesicles with Calb2$^+$ Myo7a$^+$ hair cells mimics the E18 mouse saccule (sagittal view) *in vivo*. nse, nonsensory epithelium. **h**, Tuj1$^+$ neurons extending processes to hair cells. **i**, The synaptic protein Snap25 is localized to the basal end of hair cells. **j**, The postsynaptic marker Syp colocalizes with Ctbp2 (arrowheads and inset). hcn, hair cell nucleus. **k**, Quantification of synapses on day 16, 20 and 24 hair cells ($n > 100$ cells, *$P < 0.05$, ***$P < 0.001$; mean ± s.d.). **l**, Overview of *in vitro* differentiation. Scale bars, 50 μm (**d, f, h**), 25 μm (**b, c, e, g**), 10 μm (**i**), 5 μm (**j**).

26. Géléoc, G. S. G., Risner, J. R. & Holt, J. R. Developmental acquisition of voltage-dependent conductances and sensory signaling in hair cells of the embryonic mouse inner ear. *J. Neurosci.* **24,** 11148–11159 (2004).
27. Oesterle, E. C., Campbell, S., Taylor, R. R., Forge, A. & Hume, C. R. Sox2 and Jagged1 expression in normal and drug-damaged adult mouse inner ear. *J. Assoc. Res. Otolaryngol.* **9,** 65–89 (2008).
28. Warchol, M. E. & Richardson, G. P. Expression of the Pax2 transcription factor is associated with vestibular phenotype in the avian inner ear. *Dev. Neurobiol.* **69,** 191–202 (2009).
29. Desai, S. S., Zeh, C. & Lysakowski, A. Comparative morphology of rodent vestibular periphery. l. Saccular and utricular maculae. *J. Neurophysiol.* **93,** 251–266 (2005).
30. Lysakowski, A. *et al.* Molecular microdomains in a sensory terminal, the vestibular calyx ending. *J. Neurosci.* **31,** 10101–10114 (2011).

METHODS

ESC culture. ESCs derived from blastocyst-stage embryos of R1 mice were maintained in feeder-free conditions using 2i-LIF medium as previously described[31]. In brief, ESCs were maintained on gelatin and used for experimentation until passage 40. N2B27 medium consisted of a 1:1 mixture of Advanced DMEM/F12 and Neurobasal medium (Invitrogen) supplemented with 1 mM GlutaMAX (Invitrogen) and 1 mM penicillin/streptomycin (STEMCELL Technologies). 2i-LIF medium was made by supplementing N2B27 medium with 1,000 U ml^{-1} leukaemia inhibitory factor (LIF; Millipore), 3 μM CHIR99021 (Stemgent) and 1 μM PD0325901 (Santa Cruz).

Days 0–3 of SFEBq differentiation were performed as described with slight modifications[32]. In brief, ESCs were dissociated with 0.25% trypsin-EDTA, resuspended in differentiation medium and plated 100 μl per well (3,000 cells) on 96-well low-cell-adhesion U-bottom plates (Lipidure Coat, NOF). Differentiation medium was G-MEM supplemented with 1.5% knockout serum replacement (Invitrogen), 0.1 mM nonessential amino acids, 1 mM sodium pyruvate, 1 mM penicillin/streptomycin and 1 mM 2-mercaptoethanol. On day 1, half of the medium in each well was exchanged for fresh differentiation medium containing Matrigel (2% (v/v) final concentration). On day 3 of the protocol, BMP4 (10 ng ml^{-1}) and SB-431542 (1 μM) were added to each well at 5× concentration in 25 μl of fresh media. On days 4–5, FGF2 (25 ng ml^{-1}) and LDN-193189 (100 nM) were added to each well at 6× concentration in 25 μl of fresh media. The concentration of Matrigel was maintained at 2% (v/v) throughout days 1–8. On day 8 of differentiation, cell aggregates were transferred to 24-well plates (Lipidure Coat, NOF; 4–8 aggregates per well) in N2 medium containing 1% (v/v) Matrigel. N2 medium contained Advanced DMEM/F12, 1× N2 Supplement, 1 mM penicillin/streptomycin or 50 μg ml^{-1} Normocin (InvivoGen) and 1 mM GlutaMAX. For some experiments small molecules were added to N2 medium before plating the aggregates. Half of the medium was changed every day during long-term floating culture for up to 30 days.

Signalling molecules and recombinant proteins. The following small molecules and recombinant proteins were used: recombinant human BMP4 (10 ng ml^{-1}; Stemgent), human FGF2 (25 ng ml^{-1}; PeproTech), XAV939 (1 μM; Santa Cruz), SU5402 (10 μM; BioVision), SB-431542 (1 μM; Tocris Bioscience) and LDN-193189 (100 nM; Stemgent). Notably, we have obtained comparable results using concentrations of up to 1 μM LDN-193189.

Quantitative PCR. As described previously[33], RNA was isolated using the RNeasy Minikit (Qiagen) and treated with TURBO DNase (Ambion). Single-stranded complementary DNA was synthesized using Omniscript reverse transcriptase (Qiagen) and Oligo-dT primers. All amplicons had standardized sizes of 100–110 base pairs. cDNA samples were amplified on an ABI PRISM 7900HT Sequence Detection System (Applied Biosystems) using the SYBR Green PCR Master Mix (Applied Biosystems). For each PCR reaction, a mixture containing cDNA template (5 ng), Master Mix, and forward and reverse primers (400 nM each) was treated with uracil N-glycosylase at 50 °C for 2 min before undergoing the following program: 1 cycle, 95 °C, 10 min; 45 cycles, 95 °C, 15 s, 60 °C, 1 min; 1 cycle, 95 °C, 15 s, 60 °C, 15 s, 95 °C, 15 s; 72 °C, hold. Melting curve analysis was performed to confirm the authenticity of the PCR product. The mRNA level for each gene was calculated relative to *L27* mRNA expression.

Primers used: *Dlx3* forward: CAGTACGGGAGCGTACCGGGA, reverse: TGCCGTTCACCATGCGAACC; *Sox1* forward: AACCAGGATCGGGTCAAG, reverse: ATCTCCGAGTTGTGCATCTT; brachyury forward: CACACGGCTGTGAGAGGTACCC, reverse: TGTCCGCATAGGTTGGAGAGCTC; *Pax8* forward: CGGCCGATGCCTCACAACTCG, reverse: TGGGCCAAGTCCACAATGCG; *Pax2* forward: CCCGTTGTGACCGGTCGTGATAT, reverse: TGGGGTTGCCTGAGAACTCGCTC.

Immunohistochemistry. Aggregates were fixed with 4% paraformaldehyde. The fixed specimens were cryoprotected with a graded treatment of 10, 20 and 30% sucrose and then embedded in tissue freezing medium. Frozen tissue blocks were sectioned into 10- or 12-μm cryosections. For immunostaining, a 3% goat or horse serum and 0.1% Triton X-100 solution was used for primary antibody incubation. An Alexa Fluor 488-conjugated anti-mouse IgG or anti-rat IgG and an Alexa Fluor 568-conjugated anti-rabbit IgG (Invitrogen) were used as secondary antibodies. A DAPI counterstain was used to visualize cellular nuclei (Vector, VectaShield). For whole-mount staining, aggregates were placed directly into blocking solution with 1% Triton X-100 following fixation. For confocal imaging and three-dimensional reconstruction experiments, following secondary antibody incubation, aggregates were cleared using ScaleA2 solution for 1–2 days followed by ScaleB4 treatment for another 2 days as described previously[34]. Microscopy was performed on a Nikon TE2000 inverted microscope or an Olympus FV1000-MPE confocal/multiphoton microscope. Three-dimensional reconstruction was performed using Voxx (custom software developed by Indiana Center for Biological Microscopy).

The following antibodies were used: anti-E-cadherin (rabbit, Abcam; mouse, BD Biosciences), anti-N-cadherin (mouse, BD Bioscience), anti-Sox1 (rabbit, Cell Signaling Technologies), anti-Nanog (rabbit, Abcam), anti-brachyury (goat, Santa Cruz Biotechnology), anti-AP2α (mouse, DHSB), anti-Pax8 (rabbit, Abcam), anti-Pax2 (rabbit, Invitrogen; mouse, Abnova), anti-Sox2 (mouse, BD Biosciences), anti-Jag1 (rabbit, LSBio), anti-p27^{kip1} (mouse, BD Biosciences), anti-myosinVIIa (rabbit, Proteus), anti-acetylated-α-tubulin (mouse, Abcam), anti-Tuj1 (mouse, Covance), anti-Calb2 (mouse, Millipore), anti-Caspr1 (mouse, NeuroMAB), anti-Caspr2 (mouse, NeuroMAB), anti-p63 (mouse, Santa Cruz Biotechnology), anti-cytokeratin 5 (rabbit, Sigma), anti-Nefl (rabbit, Millipore), anti-Brn3a (mouse, Millipore), anti-islet1 (mouse, DSHB), anti-Syp (rabbit, Invitrogen), anti-Brn3c (mouse, Santa Cruz Biotechnology), anti-Ctbp1 and anti-Ctbp2 (mouse, BD Biosciences), anti-Rab3 (mouse, BD Biosciences), anti-Snap25 (mouse, BD Biosciences), anti-Pax6 (rabbit, Abcam), anti-Pax3 (mouse, DSHB), anti-aPKC (rabbit, Santa Cruz Biotechnology), anti-laminin-B1 (rat, Abcam). For most of the antibodies, mouse embryonic tissue sections were used as positive controls. Mouse embryos were dissected from time pregnant CD-1 mice using a protocol approved by the Institutional Animal Care and Use Committee at Indiana University School of Medicine. The embryo fixation and processing procedure was identical to that used for cell aggregates.

The Alcian blue staining procedure was modified from a previously reported method[35]. In brief, cryosections were incubated in Alcian blue staining solution for 10 min and subsequently de-stained using 60% ethanol/40% acetic acid for 20 min. A final eosin stain was performed for 30 s. For Oil Red O staining, cryosections were kept in 60% isopropanol for 2 min and then placed in freshly prepared Oil Red O stain for 5 min followed by a 30-s haematoxylin stain.

Image analysis. The percentage of epithelial cells expressing Pax8 and Ecad was established by analysing serial sections of day 6 and day 8 aggregates. Data are representative of 6–8 aggregates from at least 3 separate experiments. For analysis of each aggregate, 5 cryosections were chosen at random positions along the z-axis of the aggregate. Using Nikon Elements or NIH ImageJ software, the Ecad$^+$ outer epithelium was outlined and cell counting of DAPI and Pax8$^+$ nuclei along the length of the epithelium established a percentage for each cryosection.

The apparent thickness of epithelia was determined by analysing cryosections stained with Ncad (control) or Ecad (all other conditions) on days 3–6. Data are representative of 6–8 aggregates from at least 3 separate experiments. For each aggregate, 3 serial sections were analysed. Five points along the epithelium were randomly chosen and the thickness was measured using Nikon Elements image analysis tools.

Similarly, the number of Myo7a$^+$ hair cells in each day 20 aggregate was determined by analysing 10-μm serial cryosections. Each biological sample represents the average number of hair cells counted in 4–6 cell aggregates and data are representative of the average from 3 separate experiments (15 aggregates total for each condition). Odd and even numbered cryosections were analysed separately and averaged to avoid double counting. The number of vesicles was quantified similarly, but every third section was analysed to avoid double counting and allow for analysis of 3 separate staining combinations. Vesicles with a long axis diameter larger than 30 μm were accounted for to avoid double counting.

Stereocilia heights were determined by measuring the apparent length of F-actin-labelled structures protruding from Myo7a$^+$ hair cells on day 20 and 24. Likewise, kinocilium heights were determined by measuring the apparent length of acetylated-α-tubulin-labelled protrusions from Myo7a$^+$ hair cells. Regions of interest were chosen randomly for analysis and >100 cells were analysed across 3–5 separate epithelia for the data shown in Fig. 3.

Synapses were quantified by analysing day 16, 20 and 24 aggregate sections stained for Syp and Ctbp2 using a previously described method[36]. Regions of interest were chosen randomly for analysis and >100 cells were analysed across 4–5 separate epithelia from 3 separate experiments for the data shown in Fig. 3. Confocal z-stacks were taken of Ctbp2-stained hair cells. The maximum-intensity projections were used to count the number of Ctbp2$^+$ puncta surrounding each hair cell nuclei.

Statistical analysis. Statistical significance was determined using a Student's t-test for comparison of two groups or a one-way analysis of variance followed by Tukey's post-hoc test for multiple comparisons, unless stated otherwise. All data were analysed using Prism 6 or Microsoft Excel software.

FM1-43 labelling. The presence of functional mechanosensitive channels was confirmed using a FM1-43 dye uptake assay similar to previous studies[25,37,38]. Large lumen aggregates (that is, >500-μm-long-axis diameter), identified by their translucency and spherical morphology relative to surrounding tissue, were used for these experiments. Aggregates were incubated in DMEM-F12 containing FM1-43FX (5 μM; Invitrogen) for 1 min and then washed 3× in fresh N2 medium. A faint cellular outline caused by autofluorescence was used to identify potential hair cells in the vesicle wall. In N2 medium, a 0.25-μm tungsten needle was used to puncture each vesicle in an area away from the site of potential hair cells. The punctured vesicles were incubated in DMEM-F12 containing FM1-43FX (5 μM) for 1 min with gentle rocking and then washed 3× in fresh N2 medium. Vesicles

were imaged to confirm dye uptake and immediately fixed with 4% paraformaldehyde. For some experiments, epithelia were fixed and incubated in PBS containing 1% Triton X-100 and phalloidin conjugated to Alexa Fluor 647 (Invitrogen) to confirm the identity of hair cells.

Electrophysiological recordings. On day 24 of differentiation, large lumen vesicles (>500-μm diameter) were dissected from cell aggregates following a 30-min incubation in DMEM/F12 containing dispase (STEMCELL Technologies). Epithelial regions containing hair cells were identified on the basis of a thickened morphology relative to the rest of the vesicle epithelium. Two incisions were made using tungsten needles on the opposite side of the vesicle in order to expose and flatten the hair-cell-containing epithelium. The flattened epithelium was mounted onto round glass coverslip and held in position by two wires glued to the coverslip using MDX4-4210 (Corning). The coverslip was then placed in a submersion-type slice chamber mounted on the stage of a Nikon E600FN Eclipse microscope. Electrophysiological recordings were performed under continuous perfusion of oxygenated artificial cerebrospinal fluid that contained the following (in mM): 130 NaCl, 3.5 KCl, 1.1 KH_2PO_4, 1.3 $MgCl_2$, 2.5 $CaCl_2$, 30 $NaHCO_3$, 10 glucose, pH 7.4 (320 mOsm kg^{-1}). Recording pipettes were pulled from borosilicate capillary glass (WPI) with resistances ranging from 2 to 3 MΩ. Recording pipettes were filled with a potassium gluconate-based recording solution that contained the following (in mM): 130 K-gluconate, 3 KCl, 3 $MgCl_2$, 5 phosphocreatine, 2 K-ATP, 0.2 NaGTP, 10 HEPES, pH 7.3 (290 mOsm kg^{-1}). Whole-cell access resistances were monitored throughout each experiment and ranged from 5–20 MΩ; a change of 15% was deemed acceptable.

Hair cells were identified with a 40× water-immersion objective and differential interference contrast. Only cells with hair bundles on their apical surface were chosen for recording. Positive pressure was maintained as the recording pipette was lowered into the epithelium. When the recording pipette touched the membrane, positive pressure was released and tight seal was formed. Recordings were obtained at 30 °C using solution inline heater (Warner Instruments). The cells were held at −60 mV, and data were acquired using whole-cell technique in voltage-clamp mode using a Multiclamp 700B amplifier (Molecular Devices) coupled to a Digidata 1332A board (Molecular Devices). The data were analysed using the pClamp 10.2 (Molecular Devices). All chemicals were purchased from Sigma-Aldrich.

Transmission electron microscopy. Day 24 aggregates were fixed in 2% paraformaldehyde/2% glutaraldehyde in 0.1 M phosphate buffer. After fixation the specimens were rinsed with phosphate buffered saline followed by post-fixation with 1% osmium tetroxide. Thereafter, the aggregates were dehydrated through a series of graded ethyl alcohols and embedded in Embed 812 (Electron Microscopy Sciences). Ultra-thin sections (70–80 nm) were cut, stained with uranyl acetate and viewed on a Tecnai BioTwin (FEI) transmission electron microscope at 80 kV. Digital images were taken with an Advanced Microscope Techniques couple-charged device camera.

Western blot analysis. Cell aggregates were lysed in radioimmunoprecipitation assay buffer supplemented with a protease inhibitor cocktail (Roche). Cell extracts were centrifuged at 13,000 r.p.m., 4 °C for 10 min to remove insoluble debris and chromosomal DNA. Proteins were separated by denaturing SDS–PAGE and transferred to PVDF membranes (Biorad). After blocking, membranes were incubated with a primary antibody overnight at 4 °C. An anti-β-actin (Sigma) antibody was used for confirmation of equal loading of the samples. Blots were detected with an horseradish peroxidase-conjugated goat anti–rabbit or rabbit anti-mouse antibody (Invitrogen) and visualized with the SuperSignal West Pico or -Femto chemiluminescent detection system (Pierce) and exposed to X-ray film.

31. Ying, Q.-L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453,** 519–523 (2008).
32. Eiraku, M. & Sasai, Y. Mouse embryonic stem cell culture for generation of three-dimensional retinal and cortical tissues. *Nature Protocols* **7,** 69–79 (2012).
33. Koehler, K. R. *et al.* Extended passaging increases the efficiency of neural differentiation from induced pluripotent stem cells. *BMC Neurosci.* **12,** 82 (2011).
34. Hama, H. *et al. Scale*: a chemical approach for fluorescence imaging and reconstruction of transparent mouse brain. *Nature Neurosci.* **14,** 1481–1488 (2011).
35. Jegalian, B. G. & De Robertis, E. M. Homeotic transformations in the mouse induced by overexpression of a human *Hox3.3* transgene. *Cell* **71,** 901–910 (1992).
36. Coate, T. M. *et al.* Otic mesenchyme cells regulate spiral ganglion axon fasciculation through a Pou3f4/EphA4 signaling pathway. *Neuron* **73,** 49–63 (2012).
37. Gale, J. E., Marcotti, W., Kennedy, H. J., Kros, C. J. & Richardson, G. P. FM1-43 dye behaves as a permeant blocker of the hair-cell mechanotransducer channel. *J. Neurosci.* **21,** 7013–7025 (2001).
38. Hu, Z. & Corwin, J. T. Inner ear hair cells produced in vitro by a mesenchymal-to-epithelial transition. *Proc. Natl Acad. Sci.* **104,** 16675–16680 (2007).

LETTER

Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells

Kathryn Blaschke[1]*, Kevin T. Ebata[1]*, Mohammad M. Karimi[2,3], Jorge A. Zepeda-Martínez[4], Preeti Goyal[2], Sahasransu Mahapatra[4], Angela Tam[3], Diana J. Laird[1], Martin Hirst[3,5], Anjana Rao[4], Matthew C. Lorincz[2] & Miguel Ramalho-Santos[1]

DNA methylation is a heritable epigenetic modification involved in gene silencing, imprinting, and the suppression of retrotransposons[1]. Global DNA demethylation occurs in the early embryo and the germ line[2,3], and may be mediated by Tet (ten eleven translocation) enzymes[4–6], which convert 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC)[7]. Tet enzymes have been studied extensively in mouse embryonic stem (ES) cells[8–12], which are generally cultured in the absence of vitamin C, a potential cofactor for Fe(II) 2-oxoglutarate dioxygenase enzymes such as Tet enzymes. Here we report that addition of vitamin C to mouse ES cells promotes Tet activity, leading to a rapid and global increase in 5hmC. This is followed by DNA demethylation of many gene promoters and upregulation of demethylated germline genes. Tet1 binding is enriched near the transcription start site of genes affected by vitamin C treatment. Importantly, vitamin C, but not other antioxidants, enhances the activity of recombinant Tet1 in a biochemical assay, and the vitamin-C-induced changes in 5hmC and 5mC are entirely suppressed in Tet1 and Tet2 double knockout ES cells. Vitamin C has a stronger effect on regions that gain methylation in cultured ES cells compared to blastocysts, and in vivo are methylated only after implantation. In contrast, imprinted regions and intracisternal A particle retroelements, which are resistant to demethylation in the early embryo[2,13], are resistant to vitamin-C-induced DNA demethylation. Collectively, the results of this study establish vitamin C as a direct regulator of Tet activity and DNA methylation fidelity in ES cells.

ES cells are derived from the inner cell mass (ICM) of the blastocyst and can be cultured in vitro to maintain a pluripotent state. Media composition has been shown previously to influence ES-cell heterogeneity, gene expression and epigenetic patterns[14]. Our study began with the serendipitous observation that culture of mouse ES cells in knockout serum replacement (KSR) strongly and reversibly induces expression of the germline gene Dazl (Supplementary Fig. 1a, b). We performed a small-molecule screen and identified vitamin C as the KSR component responsible for Dazl induction (Supplementary Fig. 2). Dazl induction was also observed with the DNA methyltransferase (Dnmt) inhibitor 5-azacytidine, suggesting that vitamin C may promote DNA demethylation. Vitamin C enhances the activity of some Fe(II) 2-oxoglutarate dioxygenases[15], and we therefore reasoned that vitamin C could promote the activity of Tet enzymes, leading to DNA demethylation and Dazl induction. As mouse ES cells are commonly cultured without vitamin C, we set out to test the effect of vitamin C on the epigenetic and transcriptional state of ES cells.

Vitamin C treatment of naive ES cells cultured with MEK and GSK3β inhibitors (2i) in N2B27-based medium, which is devoid of detectable vitamin C (Supplementary Fig. 3), leads to a striking global increase in 5hmC by immunofluorescence and dot blot (Fig. 1a, b). In contrast, global levels of 5mC were not altered at 12 or 72 h after the start of vitamin C treatment (Fig. 1b). To assess dynamic 5hmC and 5mC changes at specific genomic regions, 5hmC and 5mC DNA immunoprecipitation followed by deep sequencing (DIP-seq) was performed at 12 and 72 h after vitamin C treatment. We restricted our analysis to methylated regions, as 5mC is a prerequisite for 5hmC.

Notably, most methylated promoters transiently gain 5hmC at 12 h and return to baseline levels or below at 72 h, whereas 5mC is lost progressively at 12 and 72 h (Fig. 1c, Supplementary Figs 4a and 5). Reduction of 5hmC at 72 h may be explained by loss of 5mC substrate. After 72 h of vitamin C treatment methylation is reduced by twofold or more in 61% of analysed promoters (Supplementary Table 1). Demethylation at exons, introns and intergenic regions is also observed (Supplementary Fig. 4b). There is a highly significant overlap in the promoters that gain 5hmC at 12 h and those that lose 5mC at 72 h ($P < 2.2 \times 10^{-16}$, Fig. 1d and Supplementary Fig. 6). In addition, 5hmC gain and 5mC loss occur at the same genomic locations near the transcription start site (TSS) (example in Fig. 1e). These results support a kinetic model in which oxidation of 5mC to 5hmC precedes DNA demethylation that may occur through active or passive mechanisms. We confirmed demethylation at the promoters of three representative genes by bisulphite sequencing (Fig. 1f). Several high-density CpG promoters show minimal demethylation and many of these were identified as imprinted genes (Fig. 1g), indicating that certain regions of the genome are resistant to vitamin-C-induced demethylation.

Although dot blot analysis indicates that the global rise in 5hmC is sustained at 72 h (Fig. 1b), DIP-seq indicates that promoters return to baseline 5hmC levels at this time (Fig. 1c). This apparent discrepancy may be explained by prolonged retention of 5hmC on repetitive elements, which cover a large portion of the genome. Indeed, we find that intracisternal A particle (IAP) endogenous retroviruses (ERVs) gain 5hmC at 12 h and maintain elevated levels after 72 h of vitamin C treatment (Fig. 1h). IAP retroelements are also resistant to vitamin-C-induced demethylation, as are other repetitive elements, which may explain the maintenance of 5mC observed in the dot blot (Fig. 1h, Supplementary Fig. 7a, b and Supplementary Table 2). The increase in 5hmC at IAP retroelements does not correspond to a loss in 5mC, which could occur if only a small fraction of methylated CpGs within these ERVs gain 5hmC, resulting in no detectable loss of overall methylation by DIP. Indeed, bisulphite sequencing reveals that IAP retroelements are not demethylated with vitamin C treatment at 72 h (Supplementary Fig. 7c).

The effects of vitamin C are specific, as several other antioxidants tested did not increase global 5hmC (Supplementary Fig. 8). The effects of vitamin C are also reversible. The global increase in 5hmC

[1]Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, Department of Obstetrics and Gynecology and Center for Reproductive Sciences, University of California San Francisco, 35 Medical Center Way, San Francisco, California 94143, USA. [2]Department of Medical Genetics, Life Sciences Institute, The University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada. [3]BC Cancer Agency, Canada's Michael Smith Genome Sciences Centre, 675 West 10th Avenue, Vancouver, British Columbia V5Z 1L3, Canada. [4]La Jolla Institute for Allergy and Immunology and Sanford Consortium for Regenerative Medicine, La Jolla, California 92037, USA. [5]Department of Microbiology and Immunology, Life Sciences Institute, The University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada.
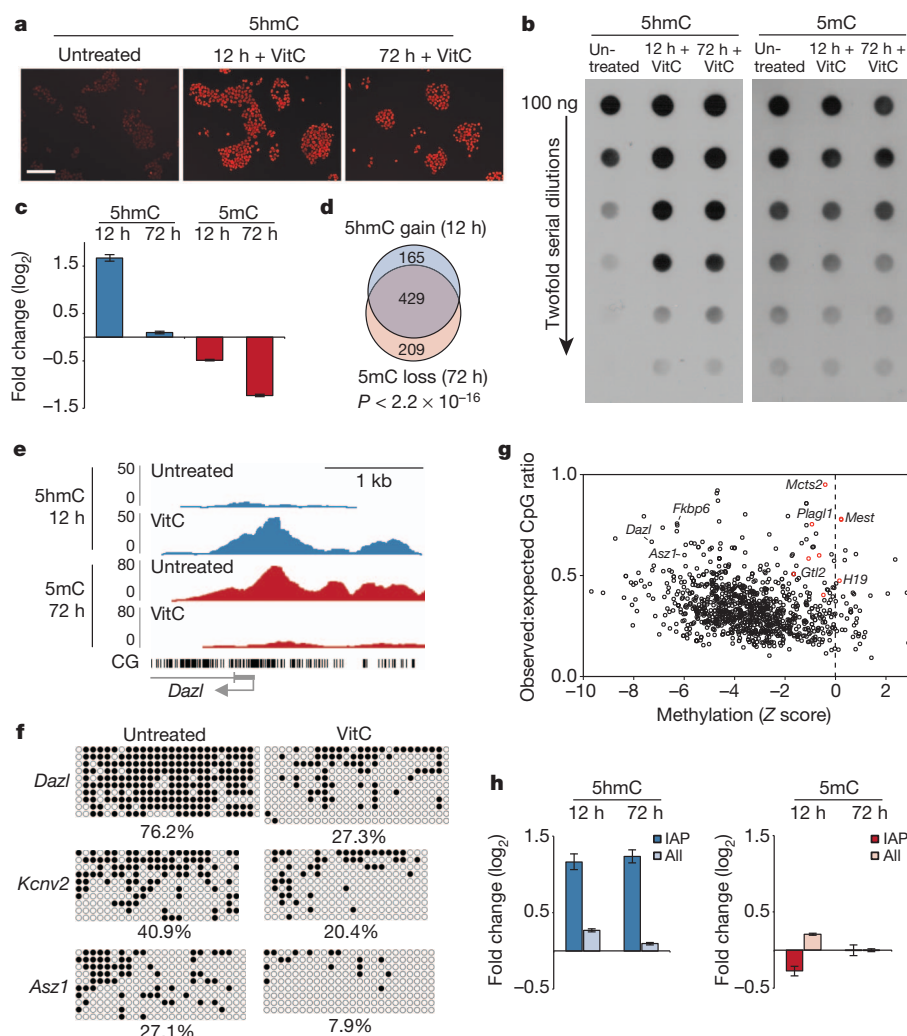*These authors contributed equally to this work.

Figure 1 | Vitamin C induces loss of 5mC at gene promoters through a transient increase in 5hmC. a, Immunofluorescence for 5hmC. Scale bar, 200 μm. VitC, vitamin C. **b**, Global 5hmC and 5mC levels assayed by dot blot analysis. **c**, Graph shows fold change in DIP-seq reads (reads per kilobase per million, RPKM) at methylated promoters ($n = 1,045$) for vitamin-C-treated cells relative to untreated cells. Values are mean ± s.e.m. **d**, Overlap of methylated promoters that gain 5hmC and those that lose 5mC. P value was calculated using Fisher's exact test. **e**, Genome browser view of *Dazl*. **f**, Bisulphite sequencing of promoters. Open circles, unmethylated; closed circles, methylated. **g**, Scatter plot of methylated promoters comparing change in methylation (Z score) with CpG content. Red circles, imprinted genes. **h**, Graphs show fold change in RPKM at retrotransposons for vitamin-C-treated cells relative to untreated cells. Values are mean ± s.e.m.

is lost rapidly after 3 days of vitamin C withdrawal, whereas promoter 5mC increases gradually after vitamin C removal (Supplementary Fig. 9).

To determine how vitamin C affects gene expression in ES cells, microarray experiments were performed. Only approximately 200 genes are changed by more than twofold, and most are upregulated (Fig. 2a and Supplementary Table 3), consistent with loss of a silencing mark like 5mC. Upregulated genes are enriched on the X chromosome (32.7% observed versus 3.8% expected) and for germline gene ontology terms[16] (Fig. 2b, c). Pluripotency gene expression is not affected (Supplementary Table 3) and vitamin C treatment does not impair differentiation (Supplementary Fig. 10). Importantly, the expression of Tet and Dnmt genes is not affected by vitamin C treatment (Supplementary Fig. 11).

Germline genes are also induced in ES cells lacking Dnmts, as reported previously[17,18]. Out of the 134 vitamin-C-induced genes, 48 (36%) are also upregulated in $Dnmt1^{-/-};Dnmt3a^{-/-};Dnmt3b^{-/-}$ (Dnmt triple knockout) ES cells, which are devoid of DNA methylation (Supplementary Fig. 12a, b). Notably, vitamin C further increases expression of a subset of these genes in Dnmt triple knockout ES cells (Supplementary Fig. 12b), suggesting that vitamin C may regulate gene expression by additional mechanisms. For example, vitamin C may also stimulate histone demethylases, as has been shown in induced pluripotent stem cell (iPS cell) generation[19].

Genes upregulated by vitamin C have higher basal levels of promoter methylation in untreated cells (Fig. 2d). When analysis is restricted to germline genes (associated with the gene ontology term 'reproduction'), genes upregulated by vitamin C show even higher basal levels of promoter methylation (Fig. 2d). Furthermore, upregulated genes, particularly upregulated germline genes, show significant loss of methylation (Fig. 2e). Taken together, these results indicate that widespread promoter demethylation induced by vitamin C promotes the upregulation of predominantly germ line-associated genes. However, promoter demethylation is not sufficient to induce expression of most methylated genes, possibly owing to redundant epigenetic silencing mechanisms or a lack of activating transcription factors.

As Tet1, Tet2 and Tet3 are the only known enzymes that oxidize 5mC to 5hmC, we reasoned that the effects of vitamin C would be mediated by Tet enzymes. Indeed, vitamin C, but not other antioxidants such as glutathione or dithiothreitol (DTT), increases recombinant Tet1 activity in a dose-dependent manner in a biochemical assay (Fig. 3a). ES cells express two Tet family members, Tet1 and Tet2, and these enzymes seem to be highly redundant[20,21]. Tet1 binding[10] is enriched near the TSS of promoters that gain 5hmC or lose 5mC with vitamin C treatment (Supplementary Fig. 13). To test whether the effects of vitamin C are Tet-dependent, we analysed $Tet1^{-/-};Tet2^{-/-}$
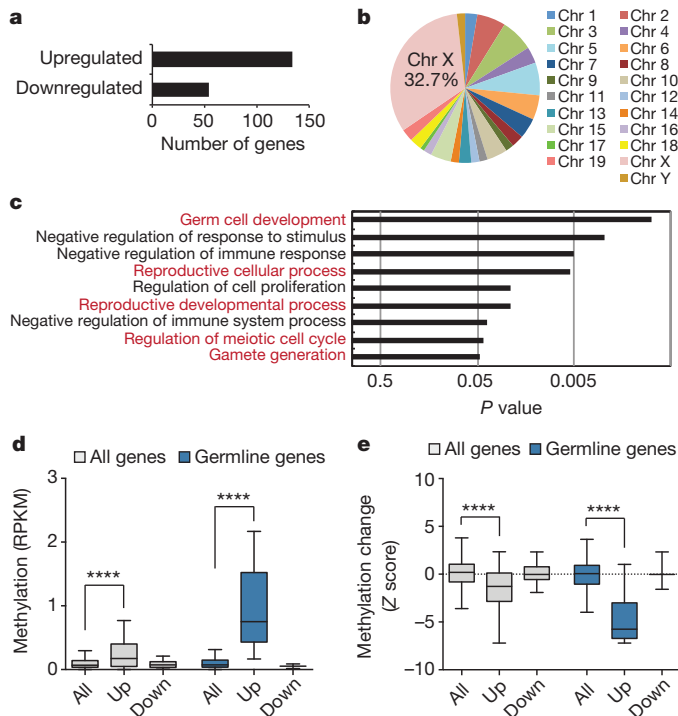
Figure 2 | Vitamin-C-induced DNA demethylation leads to expression of germline genes. a, Number of genes differentially expressed after vitamin C treatment (twofold and $P < 0.05$ by t-test). **b**, Chromosomal distribution of upregulated genes. **c**, Gene ontology analysis of upregulated genes. **d**, Box plot showing basal promoter methylation levels (RPKM) in untreated ES cells for all genes on the microarray ($n = 18,023$), upregulated genes ($n = 102$), downregulated genes ($n = 48$), all germline genes ($n = 865$), upregulated germline genes ($n = 8$), and downregulated germline genes ($n = 3$). **e**, Box plot showing the extent of vitamin-C-induced demethylation (Z score) at gene promoters categorized as in **d**. The box plots have Tukey whiskers, a line for the median, and edges for the 25th and 75th percentiles. ****$P < 0.0001$ by analysis of variance (ANOVA) throughout the figure.

Figure 3 | The effects of vitamin C are Tet-dependent. a, Dose-dependent effect of vitamin C on *in vitro* Tet activity ($n = 3$ technical replicates, values are mean ± s.d.). **b**, Dot blot analysis for 5hmC after 12 h vitamin C treatment. DKO, double knockout; WT, wild-type. **c**, 5hmC (top) and 5mC (bottom) DIP-qPCR after 12 or 72 h vitamin C treatment, respectively. An intergenic region on chromosome 8 (Int8) is included as a negative control. **d**, Gene expression at 72 h of vitamin C treatment. Nanog, whose expression is not expected to change, is included as a control. qRT–PCR data expressed relative to untreated wild-type cells. For **c** and **d**, $n = 2$ biological replicates, values are mean ± s.e.m. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ by t-test throughout the figure.

(*Tet* double knockout) ES cells[21]. Dot blot analysis reveals that *Tet* double knockout ES cells show greatly reduced 5hmC signal that is not increased after vitamin C treatment (Fig. 3b). Residual signal in *Tet* double knockout ES cells may be due to antibody background or low-level Tet3 expression. Importantly, DIP followed by quantitative polymerase chain reaction (DIP-qPCR) reveals that in contrast to wild-type cells, vitamin C treatment of *Tet* double knockout ES cells does not affect 5hmC or 5mC levels at gene promoters (Fig. 3c). Furthermore, vitamin-C-induced gene expression is significantly attenuated in *Tet* double knockout ES cells (Fig. 3d). The modest gene induction observed may be due to effects of vitamin C unrelated to DNA methylation, as already suggested by the analysis of *Dnmt* triple knockout ES cells (Supplementary Fig. 12b). $Tet1^{-/-}$ (*Tet1* knockout) ES cells[20] also show an attenuated increase in global 5hmC, reduced promoter demethylation, and reduced gene induction in response to vitamin C. However, these effects are more subtle than in the *Tet* double knockout ES cells (Supplementary Fig. 14). These data indicate that the effects of vitamin C are Tet-dependent and are mediated by both Tet1 and Tet2.

Recent studies highlight differences in the methylomes of ES cells and blastocysts, with ES cells showing higher levels of methylation[13,22]. Using published genome-wide bisulphite sequencing data for ES cells and blastocysts[13], we investigated the relationship between vitamin-C-induced demethylation and differences in methylation between ES cells and the blastocyst. Analysis was performed on promoter CpG islands (CGIs) methylated in both our study and ref. 13 (Fig. 4a and Supplementary Table 4), which generally show greater methylation in ES cells than blastocysts (Fig. 4b). Interestingly, vitamin C induces greater demethylation at CGIs that are hypermethylated in ES cells
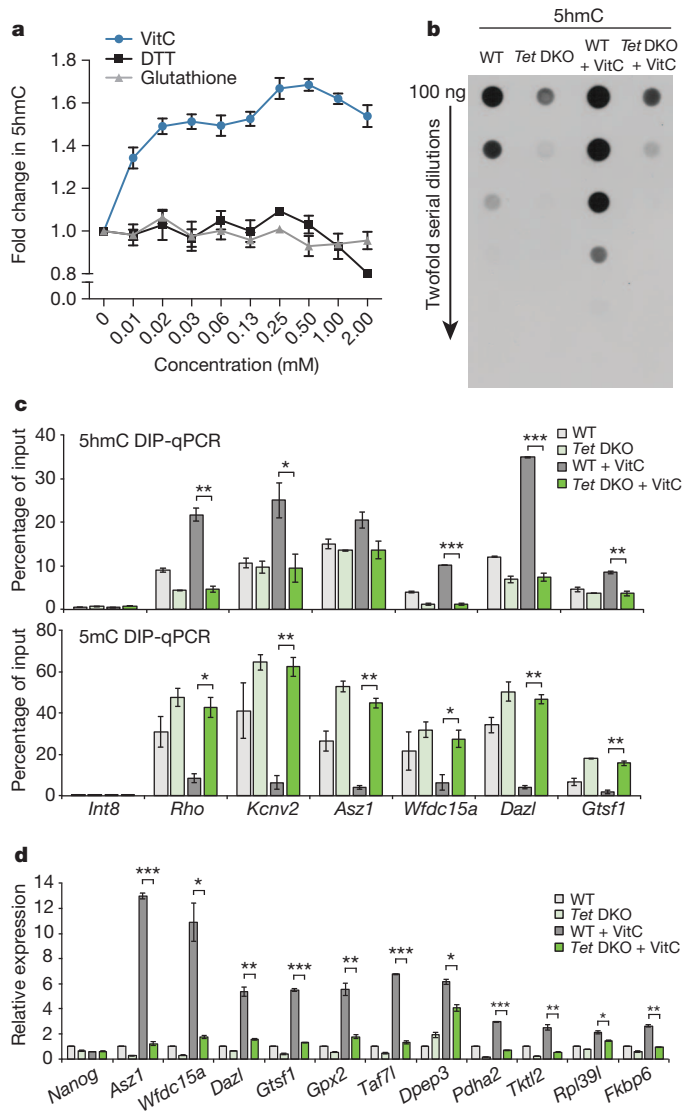
relative to blastocysts (Fig. 4c, d). Conversely, vitamin C has modest effects on CGIs with similar methylation levels in ES cells and blastocysts, such as imprinted regions (Fig. 4c, d). IAP ERVs, which are similarly methylated in both ES cells and the blastocyst, are also resistant to vitamin-C-induced demethylation in ES cells (Supplementary Fig. 7d). These findings suggest that the effects of vitamin C are most pronounced at genes that show hypermethylation in ES cells compared to blastocysts.

Next, methylation dynamics during development were determined using a reduced representation bisulphite sequencing (RRBS) data set[2]. Gene promoters that show 5mC loss with vitamin C are generally unmethylated up to the ICM stage and then undergo extensive methylation at the epiblast stage (Fig. 4e, blue). In contrast, gene promoters that show 5mC maintenance with vitamin C are enriched for germ line differentially methylated regions (DMRs) that maintain approximately 50% methylation in somatic tissues throughout development
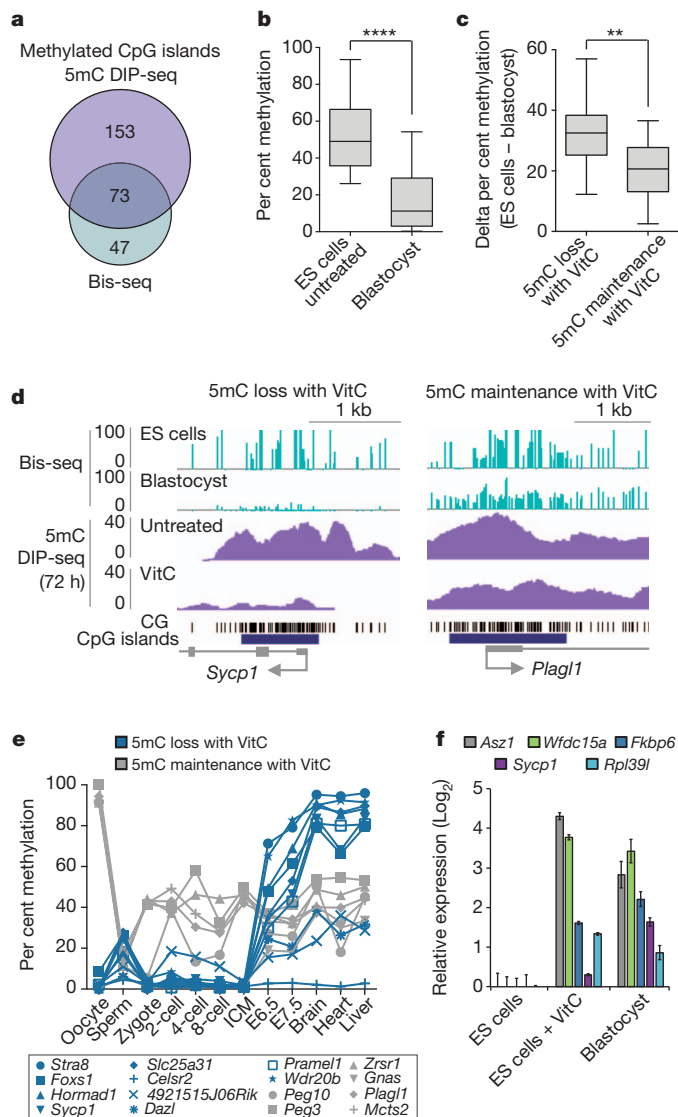
Figure 4 | Vitamin C reduces DNA methylation in ES cells that is normally gained post-implantation. **a**, Overlap of methylated CGIs in ES cells from this study (5mC DIP-seq, RPKM > 0.5) and ref. 13 (whole-genome bisulphite sequencing (Bis-seq), >25% methylated CpGs). Only CGIs found to be methylated in both data sets were used for subsequent analysis in **b–d**. **b**, The box plot shows CGI methylation levels in ES cells and blastocysts (values from ref. 13, ****$P < 0.0001$ by t-test. **c**, CGIs were first categorized as having 5mC loss with vitamin C (>75% loss of 5mC at 72 h, $n = 23$) or 5mC maintenance with vitamin C (<25% loss of 5mC at 72 h, $n = 14$) in ES cells, then plotted for difference in methylation between untreated ES cells and blastocysts. CGIs demethylated upon vitamin C treatment show significantly greater ES cell hypermethylation compared to CGIs resistant to vitamin C (**$P < 0.01$ by t-test). **d**, Genome browser views of a gene from each category described in **c**. **e**, Methylation levels during development of all genes categorized in **c** for which data exist (see ref. 2). **f**, Gene expression in ES cells cultured with or without vitamin C compared to embryonic day 3.5 (E3.5) blastocysts. Data are expressed relative to untreated ES cells ($n = 2$ biological replicates, values are mean ± s.e.m).

(Fig. 4e, grey). Thus, it seems that in ES cells cultured in the absence of vitamin C, methylation accumulates at the subset of CGIs that would normally be *de novo* methylated in the epiblast. Analysis of published RNA-seq data[23] indicates that several vitamin-C-induced genes are also expressed in the ICM (Supplementary Table 5). Furthermore, we find that several of these germline genes are indeed expressed in the blastocyst at levels comparable to ES cells treated with vitamin C (Fig. 4f). Collectively, these findings suggest that vitamin C remodels DNA methylation and expression patterns in ES cells, resulting in a state reminiscent of the ICM of the blastocyst.

Recently, it was reported that long-term culture of ES cells in 2i medium induces a blastocyst-like state of global hypomethylation relative to culture in serum[24]. As the analyses described above were carried out in 2i medium, they document effects of vitamin C beyond those of 2i. Nevertheless, we sought to distinguish the effects of 2i and vitamin C. We find that vitamin C induces a gain of 5hmC, loss of 5mC, and induction of germline genes in both FBS and 2i medium (Supplementary Fig. 15a–d). In contrast, culture in 2i medium alone shows little to no effect over the same 72-h time course. The faster kinetics of action of vitamin C relative to 2i are probably due to their different mechanisms of action: vitamin C promotes Tet-mediated DNA demethylation, whereas 2i promotes passive loss of DNA methylation via upregulation of Prdm14 and repression of Dnmt3b and Dnmt3l (ref. 24) (Supplementary Fig. 15e).

In summary, this work demonstrates that vitamin C alters the steady-state of DNA methylation and, in turn, the expression of germline genes in ES cells by enhancing Tet activity. Vitamin C reduces methylation at CGIs that normally gain methylation during the blastocyst to epiblast transition, promoting an ICM-like DNA methylation state in ES cells. Intriguingly, although human ES cells are normally cultured in medium containing vitamin C, they resemble mouse epiblast cells more than ICM cells. Nevertheless, vitamin C may also have a role in human ES cells, as it has been shown that they accumulate DNA methylation after several passages in the absence of vitamin C, although the underlying mechanisms were not addressed[25]. Notably, imprinted regions and IAP retroelements, which are resistant to DNA demethylation in the early embryo[2,13], are also resistant to vitamin-C-mediated demethylation. These regions show high levels of H3K9me3 (ref. 18), suggesting that this mark, or readers of this mark, may have a role in protecting against Tet-mediated demethylation. Vitamin C also improves the quality of iPS cells by preserving the fidelity of DNA methylation at imprinted regions[26]. Furthermore, Tets are required for methylation reprogramming during iPS cell generation and in the zygote[4,27,28]. Vitamin C has also been reported to increase 5hmC in mouse embryonic fibroblasts[29], suggesting that the mechanism characterized here may be broadly applicable to other cell types. Much work remains to be carried out to evaluate the ability of vitamin C to modulate Tet activity and DNA methylation *in vivo*. It will be of interest to investigate the role of vitamin C in contexts in which Tet enzymes have been implicated[30], such as in the zygote, germ line, blood and brain. Potential roles for vitamin C in the clinic, including in *in vitro* fertilization culture medium or in cancers driven by aberrant DNA methylation, also deserve exploration.

METHODS SUMMARY

ES cells were cultured in feeder-free conditions in 2i medium. Vitamin C (L-ascorbic acid 2-phosphate, Sigma) was added daily at 100 μg ml^{-1}. The Tet activity assay was performed using recombinant human Tet1 catalytic domain with 5hmC generation quantified by enzyme-linked immunosorbent assay (ELISA). Dot blot analyses were performed with serial dilutions of DNA using a Bio-Dot (Bio-Rad) apparatus and antibodies against 5hmC or 5mC. For gene expression analysis, total RNA was isolated and hybridized to Affymetrix mouse gene 1.0 ST GeneChip arrays or analysed by qPCR with reverse transcription (qRT–PCR). For 5mC DIP-seq and 5hmC DIP-seq, immunoprecipitated DNA was adaptor-ligated for paired-end sequencing on an Illumina HiSeq following the manufacturer's recommended protocol. Sequence reads were aligned to the mm9 mouse reference genome and unique reads were used to calculate RPKM values in various regions including RefSeq promoters and CGIs. For pair-wise sample comparisons, an empirical Z score was calculated assuming the distribution of RPKMs for each sample followed a Poisson model.

Full Methods and any associated references are available in the online version of the paper.

1. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).

2. Smith, Z. D. *et al.* A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484**, 339–344 (2012).
3. Seisenberger, S. *et al.* The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol. Cell* **48**, 849–862 (2012).
4. Gu, T.-P. *et al.* The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**, 606–610 (2011).
5. Yamaguchi, S. *et al.* Tet1 controls meiosis by regulating meiotic gene expression. *Nature* **492**, 443–447 (2012).
6. Hackett, J. A. *et al.* Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science* **339**, 448–452 (2013).
7. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
8. Pastor, W. A. *et al.* Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473**, 394–397 (2011).
9. Wu, H. *et al.* Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* **473**, 389–393 (2011).
10. Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343–348 (2011).
11. Ficz, G. *et al.* Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**, 398–402 (2011).
12. Koh, K. P. *et al.* Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell* **8**, 200–213 (2011).
13. Kobayashi, H. *et al.* Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS Genet.* **8**, e1002440 (2012).
14. Marks, H. *et al.* The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**, 590–604 (2012).
15. Loenarz, C. & Schofield, C. J. Expanding chemical biology of 2-oxoglutarate oxygenases. *Nature Chem. Biol.* **4**, 152–156 (2008).
16. Wang, P. J., McCarrey, J. R., Yang, F. & Page, D. C. An abundance of X-linked genes expressed in spermatogonia. *Nature Genet.* **27**, 422–426 (2001).
17. Fouse, S. D. *et al.* Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell* **2**, 160–169 (2008).
18. Karimi, M. M. *et al.* DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mES cells. *Cell Stem Cell* **8**, 676–687 (2011).
19. Wang, T. *et al.* The histone demethylases Jhdm1a/1b enhance somatic cell reprogramming in a vitamin-C-dependent manner. *Cell Stem Cell* **9**, 575–587 (2011).
20. Dawlaty, M. M. *et al.* Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell* **9**, 166–175 (2011).
21. Dawlaty, M. M. *et al.* Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. *Dev. Cell* **24**, 310–323 (2013).
22. Borgel, J. *et al.* Targets and dynamics of promoter DNA methylation during early mouse development. *Nature Genet.* **42**, 1093–1100 (2010).
23. Tang, F. *et al.* Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. *Cell Stem Cell* **6**, 468–478 (2010).
24. Leitch, H. G. *et al.* Naive pluripotency is associated with global DNA hypomethylation. *Nature Struct. Mol. Biol.* **20**, 311–316 (2013).
25. Chung, T.-L. *et al.* Vitamin C promotes widespread yet specific DNA demethylation of the epigenome in human embryonic stem cells. *Stem Cells* **28**, 1848–1855 (2010).
26. Stadtfeld, M. *et al.* Ascorbic acid prevents loss of Dlk1-Dio3 imprinting and facilitates generation of all-iPS cell mice from terminally differentiated B cells. *Nature Genet.* **44**, 398–405 (2012).
27. Doege, C. A. *et al.* Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. *Nature* **488**, 652–655 (2012).
28. Gao, Y. *et al.* Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. *Cell Stem Cell* **12**, 453–469 (2013).
29. Minor, E. A., Court, B. L., Young, J. I. & Wang, G. Ascorbate induces Ten-eleven translocation (Tet) methylcytosine dioxygenase-mediated generation of 5-hydroxymethylcytosine. *J. Biol. Chem.* **288**, 13669–13674 (2013).
30. Tan, L. & Shi, Y. G. Tet family proteins and 5-hydroxymethylcytosine in development and disease. *Development* **139**, 1895–1902 (2012).

METHODS

Cell culture. ES cells used in this study include: *Oct4*-GiP (129 × MF1), V6.5 (129/Sv × C57/BL6), *Tet1*$^{-/-}$, *Tet1*$^{-/-}$ *Tet2*$^{-/-}$, J1 (129/Sv), and *Dnmt1*$^{-/-}$ *Dnmt3a*$^{-/-}$ *Dnmt3b*$^{-/-}$. All ES cells were cultured in feeder-free conditions on 0.1% gelatin-coated tissue culture plates. Unless indicated otherwise, ES cells were cultured in 2i medium, which is composed of a N2B27 base medium[31] supplemented with the MEK (MAP-kinase kinase) inhibitor, PD0325901 (1 μM, Stemgent), the GSK3β inhibitor, CHIR99021 (3 μM, Selleck Chemicals), and with ESGRO leuk-aemia inhibitory factor (LIF) at 1,000 units per ml (Millipore). Vitamin C (L-ascorbic acid 2-phosphate, Sigma, A8960) was added on day 1 after seeding at 100 μg ml^{-1}. Medium was replaced daily. For KSR versus FBS studies, a base medium composed of high glucose DMEM (Life Technologies), L-glutamine (2 mM, Life Technologies), sodium pyruvate (1 mM, Life Technologies), non-essential amino acids (1×, Life Technologies), 2-mercaptoethanol (1×, Millipore), and penicillin-streptomycin (1×, Life Technologies) was supplemented with LIF (1,000 units per ml) and either 15% KSR (KSR medium) or 15% FBS (FBS medium). For differentiation experiments, *Oct4*-GiP ES cells maintained in 2i medium were treated with vitamin C for 72 h. Untreated and vitamin-C-treated ES cells were then transferred to 60-mm Petri dishes and cultured in suspension in FBS medium without LIF to induce embryoid body formation. Medium was replaced on day 3 and embryoid bodies were collected on day 5 of differentiation. To compare the effects of vitamin C in FBS medium versus 2i medium, J1 ES cells maintained in FBS medium were switched to FBS medium plus vitamin C, 2i medium, or 2i medium plus vitamin C and collected at 12 and 72 h after changing conditions. Media were replaced daily.

Small molecule screen. *Oct4*-GiP ES cells were cultured in FBS medium. Small molecules were added at the time of seeding. The small molecules used included valproic acid (2 mM, Calbiochem), trichostatin A (20 nM, Sigma), PD0325901 (1 μM, Stemgent), CHIR99021 (3 μM, Selleck Chemicals), forskolin (10 μM, Sigma), A-83-01 (1 μM, Stemgent), RG108 (250 nM, Stemgent), 5-azacytidine (1 μM, Sigma), BIX01294 (0.5 μM, Stemgent), UNC0638 (0.5 μM, Sigma), insulin (5 μg ml^{-1}, Sigma), vitamin C (100 μg ml^{-1}, Sigma) and the lipid cocktail Albumax II (5 mg ml^{-1}, Life Technologies). The KSR components tested include vitamin C, Albumax II, and insulin (international patent application WO 98/30679).

Vitamin C quantification assay. The amount of vitamin C in 2i medium was determined using an Ascorbic Acid Assay Kit (Abcam, ab65656). Known amounts of vitamin C (L-ascorbic acid, Sigma, A4544) were added to the medium and tested as controls.

Antioxidant treatment. *Oct4*-GiP ES cells were cultured in 2i medium and treated for 24 h with antioxidants. The antioxidants tested included a modified glu-tathione, glutathione reduced ethyl ester (GMEE), at 1.5 μg ml^{-1}, sodium selenite (20 nM), vitamin B1 (9 μg ml^{-1}), vitamin E (25 μM), L-carnitine hydrochloride (15 μg ml^{-1}), and α-lipoic acid (5 μg ml^{-1}). All reagents were purchased from Sigma. After treatment, DNA was isolated and assayed by dot blot analysis.

Immunofluorescence staining. Cells were fixed in 4% paraformaldehyde for 15 min. After washing three times with PBS, cells were blocked with 5% FBS in PBST (PBS + 0.5% Tween 20) for 2 h at room temperature (20–25 °C). Primary antibodies were diluted in blocking solution and incubated with cells overnight at 4 °C. Cells were then washed three times in PBS and incubated for 2 h at room temperature with secondary antibodies diluted in blocking solution. Cells were washed three times in PBS before imaging. Primary antibodies included DAZL (1:200, Abcam) and 5-hydroxymethylcytosine (1:100, Active Motif). 594-conjugated chicken anti-rabbit (1:1,000, Life Technologies) was used as a secondary antibody.

Genomic DNA preparation. DNA was isolated using a Qiagen Gentra Puregene Kit or the phenol-chloroform-isoamyl alcohol method. RNase A digestion was included in the isolation procedure.

Dot blot analysis. Isolated DNA (1 μg per sample) was denatured in 0.1 M NaOH for 10 min at 95 °C. Samples were neutralized with 1 M NH$_4$OAc on ice, and then serially diluted twofold. DNA samples were spotted on a nitrocellulose membrane using a Bio-Dot apparatus (Bio-Rad). The blotted membrane was washed in 2× SSC buffer, dried at 80 °C for 5 min, and UV cross-linked at 120,000 μJ cm^{-2}. The membrane was then blocked in Odyssey buffer (Li-Cor) diluted 1:1 in PBS (Odyssey:PBS) overnight at 4 °C. Mouse anti-5-methylcytosine monoclonal antibody (Active Motif, 1:500) or rabbit anti-5-hydroxymethylcytosine polyclonal antibody (Active Motif, 1:5,000) in Odyssey:PBS was added for 3 h at room tem-perature. The membrane was washed for 10 min three times in PBS, and then incubated with either HRP-conjugated sheep anti-mouse immunoglobulin-G (IgG) (GE, 1:10,000) or HRP-conjugated goat anti-Rabbit IgG (Abcam, 1:10,000) secondary antibodies in Odyssey:PBS for 3 h at room temperature. The membrane was then washed for 10 min three times in PBS and visualized by chemiluminescence with GE ECL Plus.

qRT–PCR. Total RNA was isolated from cultured cells using Qiagen RNeasy with on-column DNase I treatment. Complementary DNA (cDNA) was generated

from 1 μg of RNA using random hexamers to prime the reaction. The cDNA was used as template for qRT–PCR. qRT–PCR was performed in combination with the KAPA SYBR Fast ABI Prism qPCR kit on an Applied BioSystems 7900HT sequence detection system. Primer sequences are listed in Supplementary Table 6. The relative amount of each gene was normalized using two housekeeping genes (*L7* (also known as *Rpl7*) and *Ubb*), unless otherwise specified.

Blastocyst expression analysis. C57BL/6 (Simonsen) female mice were injected with 7.5 IU of pregnant mare gonadotropin (HUMC-NHPP) followed by 7.5 IU human chorionic gonadotropin (Sigma) 46 h later. Primed females were mated with DBA/2 (Simonsen) male mice. Detection of a vaginal plug was defined as 0.5 days post coitum. Embryonic day 3.5 blastocysts were obtained by flushing the uterus of superovulated females with M2 medium (Sigma). Total RNA was iso-lated from collected blastocyst using a RNeasy Micro Kit (Qiagen) with on-column DNase I treatment. qRT–PCR was performed as described above. The relative amount of each gene was normalized to *L7*. All animal work was conducted in accordance with protocols approved by the Institutional Animal Care and Use Committee at the University of California, San Francisco.

Microarray analysis. Total RNA was isolated from biological triplicates of *Oct4*-GiP ES cells treated with or without vitamin C for 72 h and hybridized to Affymetrix mouse gene 1.0 ST GeneChip arrays (Affymetrix). Only annotated probes were considered for analysis. DChip software was used for gene expression statistical analysis.

Gene ontology analysis. Gene ontology functional annotation was performed in DAVID.

In vitro recombinant Tet1 activity assay. Recombinant human TET1 catalytic domain (1.2 μg) was incubated for 10 min at 37 °C with 178.2 ng of biotinylated DNA substrate in 50 mM HEPES (pH 8.0), 50 mM NaCl, 1 mM α-ketoglutarate, 3.7 μM ammonium iron (II) sulphate hexahydrate, 0.1 mg ml^{-1} BSA, 1 mM ATP and titrating concentrations of vitamin C (Sigma, A5960), DTT (Sigma, D0632) or reduced glutathione (Sigma, G6529). The reaction was quenched by adding EDTA (11 mM). Each sample was loaded into a 384-well streptavidin coated plate in triplicates (25-μl well) and left overnight at 4 °C on a rotating platform. The next morning the wells were washed three times with 90 μl of 1× Tris-buffered saline (TBS) with 0.1% Tween (TBST). The wells were then incubated with 50 μl per well of anti-5hmC antibody (Active Motif) diluted 1:5,000 in TBST with 5% milk for 1 h at room temperature on a rotating platform. Next the wells were washed three times with 90 μl of TBST, followed by incubation with 50 μl of goat anti-rabbit peroxidase diluted 1:3,000 in TBST with 5% milk for 1 h on a rotating platform. The wells were then washed three times with TBST, and 30 μl of TMB substrate reagent mix (BD Biosciences) were added to each well. The reaction proceeded in the dark for 10 to 12 min, and was then quenched with 20 μl of 25% sulphuric acid. The absorbance of each well was measured at 450 nm. A standard curve of twofold dilutions from a fully hydroxymethylated version of the DNA substrate was used to obtain a linear regression, which was used to calculate pmols of 5hmC formed in the reaction from absorbance values. The data are presented as fold change in 5hmC relative to untreated.

Bisulphite Sanger sequencing. *Oct4*-GiP ES cells treated with or without vitamin C for 72 h were analysed for promoter methylation. To analyse the methy-lation status of the selected genes, 0.2 μg of genomic DNA was subject to sodium bisulphite conversion using the EZ DNA Methylation-Gold kit (Zymo Research). Primers specific for the genes analysed (Supplementary Table 6) were employed in nested or semi-nested PCR reactions. PCR products were cloned through TA cloning using the pGEM-T easy kit (Promega) and individual inserts were sequenced using Genewiz Sequencing. Data were analysed using Quantification Tool for Methylation Analysis[32]. The percentage of methylated CpGs sequenced is presented for each set of samples.

5mC and 5hmC DIP-qPCR. DIP was performed using the Diagenode MagMeDIP or hMeDIP Kit with minor modifications. DNA was sonicated into short fragments (100 to 1,000 base pairs (bp)) with a Diagenode Bioruptor for 20 min with 15-s on, 15-s off cycles at low power. Sonicated DNA was heat denatured at 95 °C for 10 min. Sonicated DNA (1 μg) was immunoprecipitated with 1 μg of mouse anti-5-methylcytosine monoclonal antibody (Active Motif, 1 μg μl^{-1}) or 2.5 μg of mouse anti-5-hydroxymethylcytosine monoclonal antibody (Diagenode, 1 μg μl^{-1}). After a 2-h incubation at 4 °C, Magbeads (Diagenode) were added to the DNA-antibody mixture and samples were incubated at 4 °C overnight. Isolation of immunoprecipitated DNA was performed according to the kit instructions. qPCR was performed in combination with the KAPA SYBR Fast ABI Prism qPCR kit on a Applied BioSystems 7900HT Sequence Detection System. Primer sequences are listed in Supplementary Table 6.

5hmC and 5mC DIP-seq. *Oct4*-GiP ES cells were treated with vitamin C (100 μg ml^{-1}) for 12 and 72 h. For each sample, 12 μg of genomic DNA was isolated, split into three replicates of 4 μg each and sonicated to approximately 100 to 500 bp on a Covaris E210 platform (75 s, 10% duty cycle). Sheared DNA was end-repaired,

A-tailed, and ligated to custom paired-end adapters as described[33]. Ligated genomic DNA was size selected (100 to 300 bp) by 8% PAGE (Nuvex, Invitrogen) to remove unligated adapters. Replicates were pooled and subjected to qPCR using truncated PE1.0/2.0 PCR primers to assess ligation efficiency and quantified by fluorometer (Qubit, Life Technologies). Adaptor-ligated DNA was heat denatured at 95 °C for 10 min, rapidly cooled on ice, and immunoprecipitated with a mouse monoclonal anti-methylcytidine antibody (1 mg ml^{-1}, Eurogentec) or a mouse monoclonal anti-5-hydroxymethylcytidine (5hmC) antibody (1.6 mg ml^{-1}, Diagenode). Primary antibodies were added at 1 μl μg^{-1} of DNA and samples were incubated overnight at 4 °C with rocking agitation in 500 μl IP buffer (10 mM sodium phosphate buffer, pH 7.0, 140 mM NaCl, 0.05% Triton X-100). To recover the immunoabsorbed DNA fragments, 1 μl of rabbit anti-mouse IgG secondary antibody (2.5 mg ml^{-1}, Jackson ImmunoResearch) and 100 μl Protein A/G beads (Pierce Biotechnology) were added and incubated for 2 h at 4 °C with agitation. After immunoprecipitation beads were resuspended in TE buffer with 0.25% SDS and 0.25 mg ml^{-1} proteinase K for 2 h at 55 °C and then allowed to cool to room temperature. Immunoprecipitated and supernatant DNA were purified using Qiagen MinElute columns and eluted in 16 μl EB (Qiagen). Sequencing libraries were generated by 10 to 15 cycles of PCR using custom indexed paired-end Illumina PCR primers[34]. The resulting reactions were purified over Qiagen MinElute columns, after which a final size selection was performed by electrophoresis in 8% PAGE. Libraries were quality controlled by spectrophotometry and Agilent DNA Bioanalyzer analysis. An aliquot of each indexed library was pooled by sample and sequenced two per lane on an Illumina HiSeq2000 platform (2X76 nt + 7 nt) following the manufacturer's recommended protocol and V3 chemistry (Illumina).

Bioinformatics. Sequence reads (75 bp paired-end) were aligned to the mouse reference genome (mm9) using BWA v0.5.9 (ref. 35) and default parameters. Samtools[36] and Picard (http://picard.sourceforge.net/) were used to sort and mark duplicate reads respectively. Reads having identical coordinates were collapsed into a single read and reads with mapping qualities ≥ 5 passed to FindPeaks 4.01 (ref. 37) for generation of unthresholded and thresholded (false discovery rate (FDR) < 0.01) coverage wig files to be visualized in the UCSC genome browser[38]. To quantify the strength of 5hmC and 5mC DIP-seq marks, these wig files were used to calculate reads per kilobase per million reads (RPKM)[39,40] values in various regions of interest including RefSeq[41] promoters and CGIs (downloaded from http://genome.ucsc.edu on May 15, 2012). For pair-wise sample comparisons, an empirical Z score was calculated assuming the distribution of RPKMs for each sample followed a Poisson model: Z score = $(\text{RPKM}_A - \text{RPKM}_B)/\sqrt{(\text{RPKM}_A - r_{AB}\text{RPKM}_B)}$, where RPKM_A and RPKM_B are RPKMs in the region of interest of A and B samples, respectively, and $r_{AB} = N_A/N_B$, where N_X is the total number of aligned reads used for normalization. For promoter analysis, promoters were defined as ±500 bp from the TSS. Methylated promoters were defined as having >0.5 RPKM values in both 12 and 72 h untreated 5mC DIP-seq samples

($n = 1,045$). For previously published Tet1 ChIP-seq (chromatin immunoprecipitation followed by sequencing) data (ref. 10), short read data were downloaded from the NCBI GEO archive (GSE24843) and remapped to mm9 (NCBI 37). ChIP-seq reads with identical coordinates were collapsed into a single read. Tet1 reads were directionally extended by 300 bp using FindPeaks and unthresholded coverage wig files were generated to create Tet heatmaps in methylated promoters using ChAsE[42], an interactive analysis and exploration tool for epigenetic data. Tet1 ChIP-seq data was corrected by control data. To determine which retroelement subfamilies show global 5hmC and 5mC DIP-seq changes in vitamin-C-treated compared to untreated ES cells, RPKM values were calculated for all subfamilies in each 5hmC and 5mC DIP-seq data set. To generate RPKM values, we calculated the total number of reads aligned to each retroelement subfamily using both unique and multiple-aligned reads, and normalized the total number of reads by total genomic bp per subfamily and total number of reads for each data set. For whole genome bisulphite-seq analysis of CGIs, the percentage of methylation for each individual cytosine was downloaded from ref. 13 (http://www.nodai-genome.org/mouse_en.html). CGI methylation levels were calculated as the average of the percentage of methylation for all cytosines in each CGI region. Analysis was restricted to CGIs ± 1,000 bp from a TSS.

31. Ying, Q.-L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519–523 (2008).
32. Kumaki, Y., Oda, M. & Okano, M. QUMA: quantification tool for methylation analysis. *Nucleic Acids Res.* **36**, W170–W175 (2008).
33. Morin, R. D. *et al.* Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* **476**, 298–303 (2011).
34. Wiegand, K. C. *et al.* ARID1A mutations in endometriosis-associated ovarian carcinomas. *N. Engl. J. Med.* **363**, 1532–1543 (2010).
35. Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* **18**, 1851–1858 (2008).
36. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
37. Fejes, A. P. *et al.* FindPeaks 3.1: a tool for identifying areas of enrichment from massively parallel short-read sequencing technology. *Bioinformatics* **24**, 1729–1730 (2008).
38. Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S. & Karolchik, D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204–2207 (2010).
39. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**, 621–628 (2008).
40. Pepke, S., Wold, B. & Mortazavi, A. Computation for ChIP-seq and RNA-seq studies. *Nature Methods* **6**, S22–S32 (2009).
41. Pruitt, K. D. & Maglott, D. R. RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Res.* **29**, 137–140 (2001).
42. Younesy, H. *et al.* An interactive analysis and exploration tool for epigenomic data. *Computer Graphics Forum (Eurographics Conference on Visualization (EuroVis) 2013)* **32** (number 3), 92–93 (2013).

LETTER

Molecular basis of binding between novel human coronavirus MERS–CoV and its receptor CD26

Guangwen Lu[1]*, Yawei Hu[2]*, Qihui Wang[1]*, Jianxun Qi[1]*, Feng Gao[3,4]*, Yan Li[1], Yanfang Zhang[1,5], Wei Zhang[1], Yuan Yuan[1,6], Jinku Bao[4], Buchang Zhang[2], Yi Shi[7], Jinghua Yan[1] & George F. Gao[1,5,6,7,8]

The newly emergent Middle East respiratory syndrome coronavirus (MERS-CoV) can cause severe pulmonary disease in humans[1,2], representing the second example of a highly pathogenic coronavirus, the first being SARS-CoV[3]. CD26 (also known as dipeptidyl peptidase 4, DPP4) was recently identified as the cellular receptor for MERS-CoV[4]. The engagement of the MERS-CoV spike protein with CD26 mediates viral attachment to host cells and virus–cell fusion, thereby initiating infection. Here we delineate the molecular basis of this specific interaction by presenting the first crystal structures of both the free receptor binding domain (RBD) of the MERS-CoV spike protein and its complex with CD26. Furthermore, binding between the RBD and CD26 is measured using real-time surface plasmon resonance with a dissociation constant of 16.7 nM. The viral RBD is composed of a core subdomain homologous to that of the SARS-CoV spike protein, and a unique strand-dominated external receptor binding motif that recognizes blades IV and V of the CD26 β-propeller. The atomic details at the interface between the two binding entities reveal a surprising protein–protein contact mediated mainly by hydrophilic residues. Sequence alignment indicates, among betacoronaviruses, a possible structural conservation for the region homologous to the MERS-CoV RBD core, but a high variation in the external receptor binding motif region for virus-specific pathogenesis such as receptor recognition.

The recent identification of a novel coronavirus, MERS-CoV— which, as of May 15th 2013, had infected 40 patients with a total of 20 fatalities—has drawn worldwide attention as a potential cause of a future pandemic[5]. Unlike most coronaviruses circulating in humans that only cause mild respiratory illness[6], MERS-CoV possibly represents a second reported coronavirus of severely high virulence after SARS-CoV, which caused over 8,000 infection cases globally in 2003, with more than 800 deaths[3]. The clinical manifestations of MERS-CoV infection include fever, cough, acute respiratory distress syndrome and, in some cases, accompanying renal failure[1,2], and are very similar to those caused by SARS-CoV. However, the novel coronavirus diverges from SARS-CoV in genomic sequence, and is much more closely related to the bat-derived HKU4 and HKU5 coronaviruses[7,8]. Consistent with phylogenetic analysis, MERS-CoV does not use the SARS-CoV receptor, angiotensin converting enzyme 2 (ACE2), as its entry receptor[9]; rather, a recent study showed that it uses human CD26 for this purpose[4]. CD26 is the third peptidase to be identified as a functional coronavirus receptor, the others being aminopeptidase N (ANPEP, also known as APN and CD13)[10,11] and ACE2 (ref. 12).

The recognition of CD26 by MERS-CoV is mediated by virus surface spike (S) protein[4]. As with other coronaviruses, the MERS-CoV S protein would be cleaved in host cells into S1 and S2 subunits (Fig. 1a). S1 engages the receptor[4] whereas S2, with typical sequence motifs homologous to those identified as the heptad repeats in class I enveloped

viruses[13–15], should mediate membrane fusion. The exploitation of the virus–receptor interaction and thus of the intervention strategies requires an atomic delineation of the receptor-binding properties of S1. On the basis of previous studies, the receptor attachment sites of coronavirus S1 subunits might locate to either the amino-terminal (such as in murine hepatitis virus[16]) or the carboxy-terminal (such as in, for example, SARS-CoV[17] and human coronavirus NL63 (ref. 18)) domain. We therefore tested individually the binding of MERS-CoV S1 and its N- and C-terminal-domain proteins to cell-surface-expressed CD26 molecules. The receptor-binding capacity was attributed to the C-terminal amino acids 367–606 of MERS-CoV S1 (Fig. 1b). We hereby referred to this domain as RBD. The potent interaction between MERS-CoV RBD and CD26 was further demonstrated by surface plasmon resonance assays, in which CD26 binds to MERS-CoV RBD with a dissociation constant (K_d) of about 16.7 nM (K_{on}, 1.79×10^5 M^{-1} s^{-1}; K_{off}, 2.99×10^{-3} s^{-1}), but does not bind to the RBD of SARS-CoV (Fig. 1c).

We crystallized MERS-CoV RBD and solved its structure at a resolution of 2.5 Å (Supplementary Table 1). Two molecules of essentially the same structure are present in the asymmetric unit. Each molecule contains 208 consecutive density-traceable amino acids from V381 to L588. A Dali[19] search within the Protein Data Bank (PDB) revealed clear structural homology between MERS-CoV RBD and SARS-CoV RBD (PDB code, 2DD8; Z score, 15.1). We therefore divided the MERS-CoV RBD structure into two subdomains: a core and an external β-sheet, using the structure of SARS-CoV RBD as a reference. The core subdomain reveals a five-stranded antiparallel β-sheet (β1, β3, β4, β5 and β10) in the centre. The connecting helices (four α-helices: α1–4 and two 3$_{10}$-helices: η1 and η2) and two small β-strands (β2 and β11) further decorate the sheet on both sides, together forming a globular fold. Three disulphide bonds, connecting C383 to C407, C425 to C478, and C437 to C585, respectively, stabilize the core-domain structure from the interior. At the solvent-exposed side, the RBD termini are clinched adjacent to each other (Fig. 2a, b). This subdomain fold is very similar to that of the SARS-CoV RBD core (a root mean squared deviation of 2.79 Å for 76 Cα pairs). Superimposition of the two structures reveals a well-aligned centre sheet and homologous peripheral helices and strands, although several intervening loops are observed to exhibit large conformational variance (Fig. 2c).

The external subdomain of MERS-CoV RBD is mainly a β-sheet structure with three large (β6, β8 and β9) and one small (β7) strand arranged in an antiparallel manner. It is anchored to the RBD core through the β5/6, β7/8 and β9/10 intervening loops, which touch the core subdomain like a clamp at both the top and bottom positions. Two small 3$_{10}$ helices (η3 and η4) and most of the connecting loops in this subdomain locate on the interior side of the sheet, hence exposing a flat exterior sheet-face to the solvent. Residues C503 and C526 form
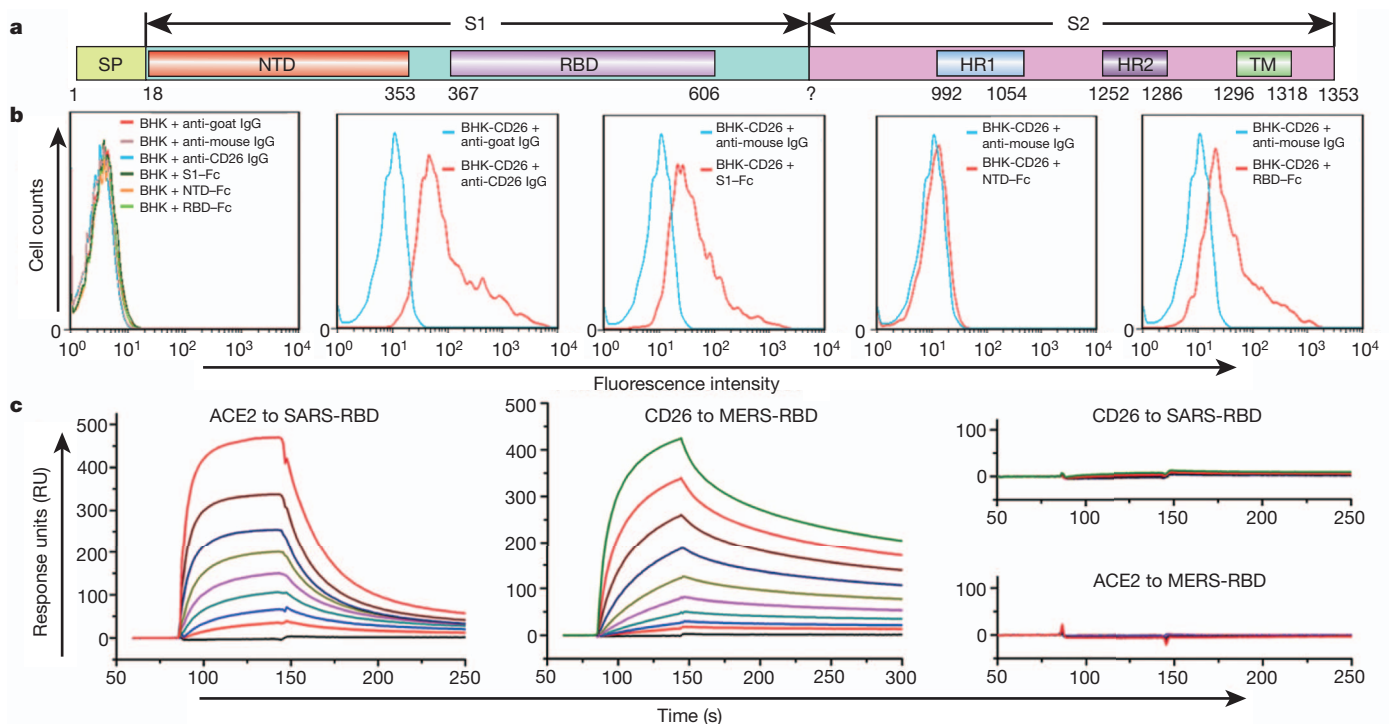
Figure 1 | Identification of the MERS-CoV RBD. a, A schematic representation of the MERS-CoV S protein. The N-terminal domain (NTD) and RBD are defined on the basis of a pairwise sequence alignment with the N-terminal galectin-like domain of murine hepatitis virus S and the RBD of SARS-CoV S, respectively. The remaining domain elements are bioinformatically defined on the basis of the web-server predictions (signal peptide (SP), SignalP 4.0 server; transmembrane domain (TM), TMHMM server; heptad repeats 1 and 2 (HR1 and HR2), Learncoil-VMF program). ? denotes the presumed/estimated S1/S2 cleavage site. A previous prediction[4] indicates cleavage between R751 and S752 with a 602-residue S2. However, a recent study[28] revealed a spike C-terminal domain (possibly S2) of ~100 kDa, indicating a cleavage site upstream of R751/S752. **b**, A flow cytometric assay of the Fc-fused S protein or its subdomain proteins involved in CD26 binding. Mock-transfected baby hamster kidney (BHK) cells or BHK cells transfected with CD26-expressing plasmid (BHK-CD26) were tested with the individual Fc-fusion proteins or an anti-CD26 antibody (anti-CD26 IgG). For each test, the secondary antibodies (anti-goat IgG or anti-mouse IgG) were used as the negative control. The profiles are shown. From left to right: BHK cells with the indicated Fc-fusion proteins or antibodies, BHK-CD26 with anti-CD26 antibody, BHK-CD26 with Fc-fused S1, BHK-CD26 with Fc-fused NTD, BHK-CD26 with Fc-fused RBD. **c**, A surface plasmon resonance assay characterizing the specific binding between CD26 and MERS-CoV RBD. The profiles are shown. Left, human ACE2 to SARS-CoV RBD; middle, CD26 to MERS-CoV RBD; top right, CD26 to SARS-CoV RBD; bottom right, human ACE2 to MERS-CoV RBD.

the fourth disulphide bond, linking the η3 helix to strand β6 (Fig. 2a, b). With no observable structure homology (Fig. 2c), the external subdomains of MERS-CoV and SARS-CoV RBDs are topological equivalents, both being present as an 'insertion' between the equivalent core-strands (strands β5 and β10 in MERS-CoV, and β6 and β9 in SARS-CoV) (Supplementary Fig. 1).

To elucidate the structural basis of the virus–receptor engagement, we further prepared the RBD–CD26 complex by *in vitro* mixture of the two proteins and then purification on a gel filtration column. Consistent with the high binding affinity between MERS-CoV RBD and CD26, the complex is easily obtainable and stable (Supplementary Fig. 2). The complex structure was solved at 2.7 Å resolution (Supplementary Table 1) with one RBD binding to a single CD26 molecule in the asymmetric unit. The receptor, as shown in previous reports[20,21], is composed of an eight-bladed β-propeller domain and an α/β hydrolase domain. MERS-CoV RBD binds to the side-surface of the CD26 β-propeller, recognizing blades IV and V and a small bulged helix in the blade-linker. As for the viral ligand, the entire receptor binding site locates in the external subdomain and to the solvent-exposed sheet-face, qualifying the subdomain as the receptor binding motif (RBM) (Fig. 3a). Overall, engagement of the receptor does not induce obvious conformational changes in RBM, although small structural variance could be observed for the tip-loops. The η2–α4 loop in the RBD core, however, unexpectedly exhibits a large conformational difference between the free and the bounded structures (Supplementary Fig. 3). We believe this is due to a crystal contact present in the free RBD structure, which is interrupted in the complex crystal by the engaging receptor.

CD26 is a type II transmembrane protein. It is present as a homodimer on the cell surface[20–22]. The dimerization of the peptidase relies on broad intermolecule contacts contributed by the hydrolase domain and the extended strands in blade IV of the β-propeller[20,21]. A lateral binding of MERS-CoV RBD to CD26 would therefore not disrupt CD26 dimerization. Accordingly, a similar U-shaped CD26 dimer could be generated by symmetry operations of the complex structure. The viral ligand locates at the membrane-distal tip of the dimer, corresponding well to a *trans* interaction between the virus and the receptor (Fig. 3b). Considering that the RBD N and C termini are on the same side distant from CD26, it is unlikely that the remaining S domains would contact the receptor molecule. The binding mode revealed by the complex structure is also in good accordance with a previous study showing that the virus–receptor interaction is independent of the peptidase activity of CD26 (ref. 4). The bound RBD is far away from interfering with either the substrate/product accessing tunnels or the catalytic centre[20,21] (Fig. 3b).

Overall, a surface area of 1203.4 and 1113.4 Å^2 in CD26 and MERS-CoV RBD, respectively, is buried by complex formation (Fig. 4a). Scrutinization of the binding interface reveals a group of hydrophilic residues at the site, forming a polar-contact (H-bond and salt-bridge) network. These interactions are predominantly mediated by the residue side chains (including RBD Y499 with CD26 R336, N501 with Q286, K502 with T288, D510 with R317, E513 with Q344, and D539 with K267), although CD26 L294 and RBD D510 are observed to contact RBD R542 and CD26 Y322, respectively, through the main-chain oxygen atom (Fig. 4b). In addition, the bulged helix in CD26
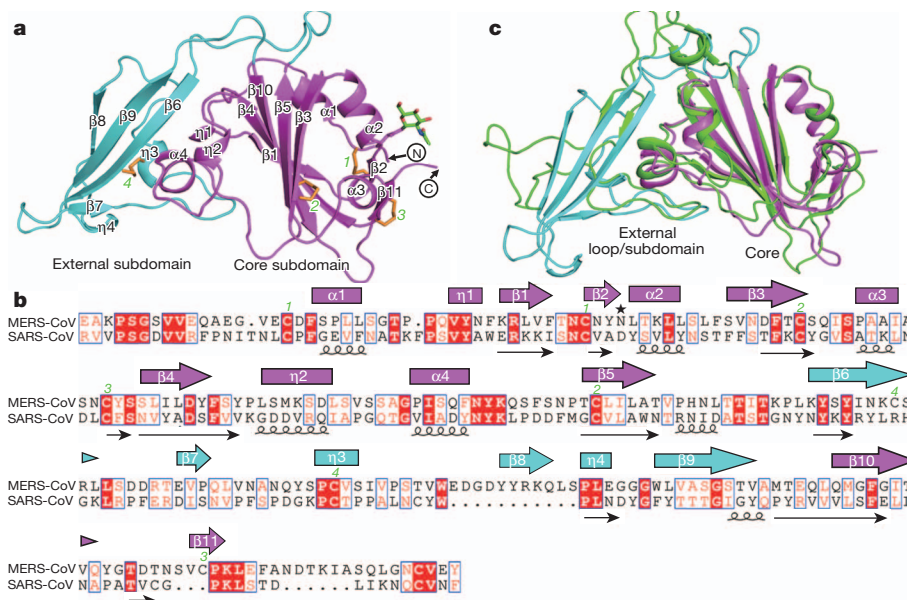
Figure 2 | The overall structure of MERS-CoV RBD. a, A cartoon representation of the RBD structure. The secondary structural elements are labelled according to their occurrence in sequence. The disulphide bonds (marked with Arabic numbers 1–4) and N-glycan linked to N410 are shown as orange and green sticks, respectively. Core subdomain, magenta; external subdomain, cyan. The N and C termini are labelled. **b**, An amino acid sequence alignment between MERS-CoV and SARS-CoV RBDs. The hollow boxes and arrows indicate α/3₁₀ helices and β-strands, respectively, and are coloured as in **a**. To facilitate comparison, the secondary-structure elements of SARS-CoV RBD (PDB code, 2DD8) are marked with spiral (helices) and arrow (strands) lines below the sequence. The cysteine residues that form disulphide bonds are labelled as in **a**, and residue N410 with a star. **c**, A structural alignment between MERS-CoV (magenta for core and cyan for external subdomains) and SARS-CoV (green) RBDs.

properly positions three hydrophobic residues A291, L294 and I295 into close proximity with the RBD amino acids Y540, W553 and V555, forming a hydrophobic centre at the interface (Fig. 4c). Further virus–receptor contacts include V341 and I346 of CD26 packing against P515 and the apolar carbon atoms of R511 and E513 in RBD (Fig. 4d), and a CD26 N229-linked carbohydrate moiety interacting with RBD amino acids W535 and E536 (Fig. 4e). Overall, the virus–receptor engagement is dominated by the polar contacts mediated by the hydrophilic residues, and mutations of those in RBD (six alanine substitutions and one Y499F mutation of the CD26-interacting amino acids) completely abrogated its interaction with CD26 (Supplementary Fig. 4). The features of these residue interactions are very similar to those mediating the interaction between adenosine deaminase (ADA) and CD26 (ref. 23). By a pairwise comparison, we unexpectedly found

that all those CD26 residues identified in the virus–receptor interface are also involved in ADA binding, indicating a competition between ADA and the virus for CD26 receptor. As the ADA–CD26 interaction is shown to induce co-stimulatory signals in T cells[22], this may indicate a possible manipulation of the host immune system by MERS-CoV through competition for the ADA-recognition site. It is also noteworthy that those CD26 residues involved in RBD binding are highly conserved between human and bat, with only two variations (I295T and R317Q), explaining the capability of MERS-CoV using bat CD26 for cell entry[4] (Supplementary Fig. 5).

Coronaviruses can be categorized into three main genera or groups (group 1 (alpha), group 2 (beta) and group 3 (gamma) coronaviruses)[24]. Both MERS-CoV and SARS-CoV belong to the betacoronavirus genus, but are classified into different lineage subgroups (subgroup 2b for
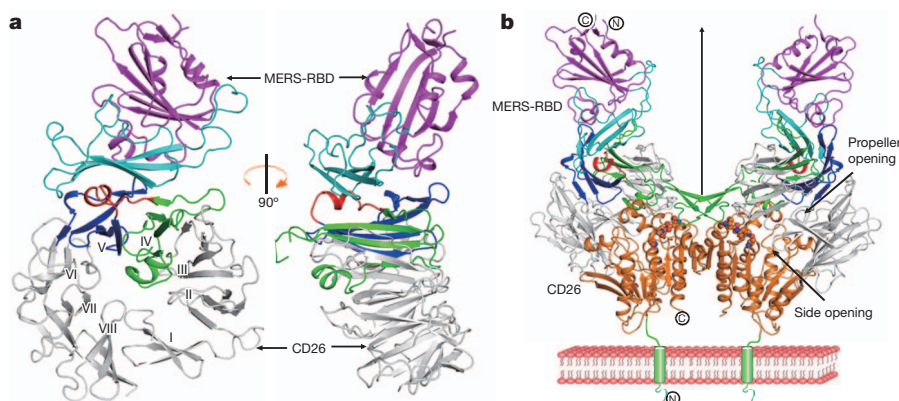

Figure 3 | The complex structure of MERS-CoV RBD bound to CD26. a, A cartoon representation of the complex structure. For clarity, only the β-propeller domain of CD26 (grey) is shown. Blades IV, V and the intervening IV/V linker that recognize RBD are highlighted in green, blue and red, respectively. The core subdomain and external RBM are coloured magenta and cyan, respectively. The right panel is yielded by clockwise rotation of the left panel along a longitudinal axis in the page-face. **b**, A symmetry-related CD26 dimer observed in the complex crystal. The two-fold axis is shown as an upright arrow. The transmembrane topology of CD26 is indicated with a modelled lipid-bilayer membrane. In CD26, the propeller and side openings indicated as the substrate entrance/exit tunnels are marked with arrows, and the catalytic triad residues are highlighted as spheres. Colour selections are the same as in **a**, and the CD26 α/β hydrolase domain is shown in orange. The N and C termini are labelled.
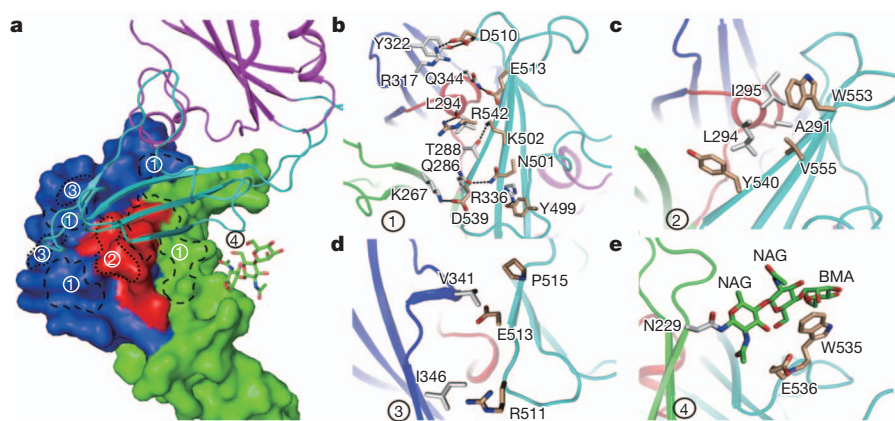
Figure 4 | The atomic interaction details at the binding interface. a, An overview of the binding interface. CD26 and RBD are shown in surface and cartoon representations, respectively, and are coloured as in Fig. 3. The carbohydrate moiety linked to CD26 N229 is shown as green sticks. The contacting sites (each allocated with an Arabic number 1–4) are further delineated in **b**–**e** for the amino acid interaction details. **b**, A strong polar-contact (H-bond and salt-bridge) network. **c, d**, The small patches of hydrophobic interactions. **e**, Contribution of the carbohydrate moiety. The residues involved are shown and labelled. NAG, *N*-acetyl-D-glucosamine; BMA, beta-D-mannose.

SARS-CoV and 2c for MERS-CoV)[8]. We noted that the spike sequences are of low identity among different subgroup members. For example, MERS-CoV and SARS-CoV S proteins show a sequence identity of less than 28%. Nevertheless, RBDs of the two coronaviruses are homologous for the core subdomain. Notably, the three interior disulphide bonds in the core are well-aligned for the steric positions in the two RBD structures and well-conserved in sequence among betacoronaviruses. Conversely, the external RBM region is highly variable in both length and residue composition (Supplementary Fig. 6). Consistently, no structural homology in this subdomain is observed between MERS-CoV and SARS-CoV. Yet it is this subdomain that engages cellular receptors. We therefore assume that betacoronaviruses probably have a similar core-domain fold in the S protein to present the external amino acids with divergent structures for viral pathogenesis, such as receptor recognition.

Our work presents the fifth structure of virus S protein–receptor complexes in the *Coronaviridae* family[16–18,25]. Taking into account both the RBD structure and the binding mode with receptors, MERS-CoV is related to SARS-CoV[17] (a single insertion functioning as RBM) but differs from porcine respiratory coronavirus[25] and NL63 (ref. 18) of alphacoronaviruses (multiple discontinuous RBMs) (Supplementary Fig. 7). Nevertheless, related structural topologies can still be observed in RBDs of these coronaviruses[26]. We noted that in the RBD–receptor complex structures of both MERS-CoV and porcine respiratory coronavirus the binding interfaces involve a receptor N-glycan. This might represent another cross-genus similarity in the *Coronaviridae* family, which supports a proposed common evolutionary origin of coronavirus S proteins[26]. It would therefore be interesting to investigate the contribution of the sugar moiety to the virus–receptor interaction for MERS-CoV in the future.

Vaccination remains the most useful measure to combat viral infection and transmission. A large number of antibodies show neutralization activity by targeting the RBD and thereby disrupting the virus–receptor engagement. Therefore, a properly folded RBD could be an ideal immunogen for vaccination, as demonstrated for SARS-CoV[27]. A recent report indeed shows the presence of S-specific neutralizing antibodies in MERS-CoV-infected patients[28]. It may be worth attempting to test the immunization effect of MERS-CoV RBD in the future.

METHODS SUMMARY

Protein expression, purification, crystallization and structure determination. Both His-tagged CD26 and MERS-CoV RBD proteins were expressed in insect High Five cells using the Bac-to-Bac baculovirus expression system (Invitrogen). The recombinant proteins were then purified via nickel-chelated affinity chromatography and gel filtration. Crystals were obtained by initial screening with the commercially available kits followed by optimization. The RBD structure was solved by single-wavelength anomalous diffraction and the complex structure by molecular replacement.

Full Methods and any associated references are available in the online version of the paper.

1. Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. & Fouchier, R. A. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* **367**, 1814–1820 (2012).
2. Bermingham, A. *et al.* Severe respiratory illness caused by a novel coronavirus, in a patient transferred to the United Kingdom from the Middle East, September 2012. *Euro Surveill.* **17**, 20290 (2012).
3. World Health Organization. Cumulative Number of Reported Probable Cases of Severe Acute Respiratory Syndrome (SARS) http://www.who.int/csr/sars/country/en/.
4. Raj, V. S. *et al.* Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* **495**, 251–254 (2013).
5. World Health Organization. Novel coronavirus infection - update. http://www.who.int/csr/don/2013_05_15_ncov/en/ (2013).
6. Weiss, S. R. & Navas-Martin, S. Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. *Microbiol. Mol. Biol. Rev.* **69**, 635–664 (2005).
7. van Boheemen, S. *et al.* Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *mBio* **3**, e00473–e12 (2012).
8. Lu, G. & Liu, D. SARS-like virus in the Middle East: a truly bat-related coronavirus causing human diseases. *Protein Cell* **3**, 803–805 (2012).
9. Müller, M. A. *et al.* Human coronavirus EMC does not require the SARS-coronavirus receptor and maintains broad replicative capability in mammalian cell lines. *mBio* **3**, e00515–e12 (2012).
10. Yeager, C. L. *et al.* Human aminopeptidase N is a receptor for human coronavirus 229E. *Nature* **357**, 420–422 (1992).
11. Delmas, B. *et al.* Aminopeptidase N is a major receptor for the entero-pathogenic coronavirus TGEV. *Nature* **357**, 417–420 (1992).
12. Li, W. *et al.* Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450–454 (2003).
13. Gao, G. F. in *Combating the Threat of Pandemic Influenza: Drug Discovery Approaches* (ed. Torrence, P.F.) 226–246 (John Wiley & Sons, 2007).
14. Skehel, J. J. & Wiley, D. C. Coiled coils in both intracellular vesicle and viral membrane fusion. *Cell* **95**, 871–874 (1998).
15. Zhu, J. *et al.* Following the rule: formation of the 6-helix bundle of the fusion core from severe acute respiratory syndrome coronavirus spike protein and identification of potent peptide inhibitors. *Biochem. Biophys. Res. Commun.* **319**, 283–288 (2004).
16. Peng, G. *et al.* Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor. *Proc. Natl Acad. Sci. USA* **108**, 10696–10701 (2011).
17. Li, F., Li, W., Farzan, M. & Harrison, S. C. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **309**, 1864–1868 (2005).
18. Wu, K., Li, W., Peng, G. & Li, F. Crystal structure of NL63 respiratory coronavirus receptor-binding domain complexed with its human receptor. *Proc. Natl Acad. Sci. USA* **106**, 19970–19974 (2009).
19. Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 (2010).

20. Rasmussen, H. B., Branner, S., Wiberg, F. C. & Wagtmann, N. Crystal structure of human dipeptidyl peptidase IV/CD26 in complex with a substrate analog. *Nature Struct. Biol.* **10**, 19–25 (2003).
21. Engel, M. *et al.* The crystal structure of dipeptidyl peptidase IV (CD26) reveals its functional regulation and enzymatic mechanism. *Proc. Natl Acad. Sci. USA* **100**, 5063–5068 (2003).
22. Gorrell, M. D., Gysbers, V. & McCaughan, G. W. CD26: a multifunctional integral membrane and secreted protein of activated lymphocytes. *Scand. J. Immunol.* **54**, 249–264 (2001).
23. Weihofen, W. A., Liu, J., Reutter, W., Saenger, W. & Fan, H. Crystal structure of CD26/dipeptidyl-peptidase IV in complex with adenosine deaminase reveals a highly amphiphilic interface. *J. Biol. Chem.* **279**, 43330–43335 (2004).
24. Lai, M. M., Perlman, S. & Anderson, L. J. in *Fields Virology* (ed. Knipe, D.M.) 1305–1336 (Lippincott Williams & Wilkins, 2007).
25. Reguera, J. *et al.* Structural bases of coronavirus attachment to host aminopeptidase N and its inhibition by neutralizing antibodies. *PLoS Pathog.* **8**, e1002859 (2012).
26. Li, F. Evidence for a common evolutionary origin of coronavirus spike protein receptor-binding subunits. *J. Virol.* **86**, 2856–2858 (2012).
27. Du, L. *et al.* The spike protein of SARS-CoV–a target for vaccine and therapeutic development. *Nature Rev. Microbiol.* **7**, 226–236 (2009).
28. Gierer, S. *et al.* The spike-protein of the emerging betacoronavirus EMC uses a novel coronavirus receptor for entry, can be activated by TMPRSS2, and is targeted by neutralizing antibodies. *J. Virol.* **87**, 5502–5511 (2013).

METHODS

Protein expression and purification. The proteins used for crystallization and surface plasmon resonance experiments were prepared with the Bac-to-Bac baculovirus expression system (Invitrogen). The coding sequences for MERS-CoV RBD (GenBank accession number JX869059, spike residues 367–606), SARS-CoV RBD (accession number NC_004718, spike residues 306–527), human CD26 (accession number NP_001926, residues 39–766) and human ACE2 (accession number BAJ21180, residues 19–615) were individually cloned into the pFastBac1 vector. For each construct, a previously described gp67 signal peptide sequence[29] was added to the protein N terminus for protein secretion, and a hexa-His tag was added to the C terminus to facilitate further purification processes. Transfection and virus amplification were conducted with Sf9 cells, and the recombinant proteins were produced in High Five cells. The cell culture was collected 48 h after infection and passed through a 5-ml HisTrap HP column (GE Healthcare). After removal of most of the impurities, the recovered proteins were then pooled and further purified on a Superdex 200 column (GE Healthcare). Finally, each collected protein was prepared in a buffer consisting of 20 mM Tris-HCl (pH 8.0) and 150 mM NaCl and concentrated to about 10 mg ml^{-1} for further use.

To obtain the complex of MERS-CoV RBD bound to CD26, the individual proteins were *in vitro* mixed at a molar ratio of 1:1 and incubated at 4 °C for about 2 h. The complex was then further purified on a Superdex 200 column, and concentrated to about 15 mg ml^{-1} for crystallization experiments.

To prepare the Fc chimaeric proteins, the fragment encoding MERS-CoV S1 (residues 1–751) or NTD (residues 1–353) or RBD (adding the S residues 1–17 of the signal peptide to its N terminus to facilitate protein secretion) was fused 5′-terminally to a fragment coding for the Fc domain of mouse IgG and ligated into the pCAGGS expression vector. A mutant RBD–Fc protein-expressing plasmid was also constructed by site-directed mutagenesis, for which the identified hydrophilic residues involved in CD26 binding were mutated simultaneously (Y499F; N501A, K502A, D510A, E513A, D539A and R542A). The expression plasmids were then transfected into HEK293T cells. The cell culture was collected 48 h after transfection and directly used in the flow cytometric assay.

Analytical gel filtration. MERS-CoV RBD, CD26 and their protein complex were individually prepared and adjusted to the same volume. The samples were then loaded onto a calibrated Superdex 200 column (GE Healthcare). The chromatographs were recorded and overlaid onto each other. The pooled proteins were analysed on a 12% SDS–PAGE gel and stained with Coomassie blue.

Surface plasmon resonance assay. The BiAcore experiments were carried out at room temperature (25 °C) using a BIAcore 3000 machine with CM5 chips (GE Healthcare). For all the measurements, an HBS-EP buffer consisting of 10 mM HEPES, pH 7.5, 150 mM NaCl, 3 mM EDTA and 0.005% (v/v) Tween-20 was used, and all proteins were exchanged to the same buffer in advance via gel filtration. The MERS-CoV RBD and SARS-CoV RBD proteins were immobilized on the chip at about 500 response units. Gradient concentrations of human CD26 (0, 5, 10, 20, 40, 80, 160, 320, 640 and 1,280 nM) or human ACE2 (0, 10, 20, 40, 80, 160, 320, 640 and 1,280 nM) were then used to flow over the chip surface. After each cycle, the sensor surface was regenerated via a short treatment using 10 mM NaOH. The binding kinetics were analysed with the software BIAevaluation Version 4.1 using the 1:1 Langmuir binding model.

Flow cytometric assay. For the surface expression of CD26, the full-length coding sequence was cloned into the pEGFP-C1 vector which yields a plasmid encoding a recombinant CD26 protein with an EGFP-tag fused to its N terminus. The plasmid was transfected into the CD26-negative BHK cells using lipo2000 (Invitrogen) according to the manufacturer's instructions. The cells were collected 48 h after transfection.

For staining, the mock-transfected BHK cells or the cells transfected with the CD26-expressing plasmid were suspended in PBS and incubated with the individual Fc-fusion protein culture or goat anti-CD26 IgG (R&D Systems) at room temperature for 1 h. The cells were then washed and further incubated at room temperature for about 0.5 h with anti-mouse or anti-goat secondary IgG antibodies (R&D Systems). After washing, the cells were analysed by flow cytometry with a BD FACSCalibur machine. The cells incubated only with the secondary antibodies were used as the negative controls.

Crystallization. All the crystals were obtained by vapour-diffusion sitting-drop method with 1 μl protein mixing with 1 μl reservoir solution and then equilibrating against 100 μl reservoir solution at 18 °C. The initial crystallization screenings were carried out using the commercially available kits. The conditions that yield crystals were then optimized. Diffractable crystals of the free RBD protein were finally obtained in a condition consisting of 0.1 M ammonium tartrate dibasic, pH 7.0, and 12% PEG 3,350 with a protein concentration of 10 mg ml^{-1}. Derivative crystals were obtained by soaking RBD crystals for 24 h in mother liquor containing 2 mM KAuCl$_4$•2H$_2$O. The complex crystals were grown in 6% (v/v) 2-propanol, 0.1 M sodium acetate pH 4.5 and 26% PEG550 with a protein concentration of 15 mg ml^{-1}.

Data collection, integration and structure determination. For data collection, all crystals were flash-cooled in liquid nitrogen after a brief soaking in reservoir solution with the addition of 20% (v/v) glycerol. The native RBD data set was collected at the High Energy Accelerator Research Organization (KEK) BL1A (wavelength, 1.03818 Å), whereas the diffraction data for the Au derivative crystal (wavelength, 1.0382 Å) and the complex crystal (wavelength, 0.97930 Å) were collected at the Shanghai Synchrotron Radiation Facility (SSRF) BL17U. All data were processed with HKL2000 (ref. 30). Additional processing was performed with programs from the CCP4 suite[31].

The structure of RBD was determined by the single-wavelength anomalous diffraction (SAD) method. The Au sites were first located by SHELXD[32] for the Au-SAD data. The identified position were then refined and the phases were calculated with SAD experimental phasing module of Phaser[33]. The real space constraints were further applied to the electron density map in DM[34]. The initial model was built with Autobuild in Phenix package[35]. Additional missing residues were added manually in Coot[36]. The final model was refined with phenix.refine in the Phenix[35] with energy minimization, isotropic ADP refinement, and bulk solvent modelling. The complex structure was solved by molecular replacement module of Phaser[33], with the solved RBD structure and previously reported CD26 structure (PDB code, 2BGR) as the search models. The atomic model was completed with Coot[36] and refined with phenix.refine[35]. The stereochemical qualities of the final models were assessed with PROCHECK[37]. The Ramachandran plot distributions for the residues in the free RBD structure were 86.8, 11.8 and 1.4% for the most favoured, additionally and generously allowed regions, respectively. These values were 86.5, 13.1 and 0.5% for the RBD–CD26 complex structure. Data collection and refinement statistics are summarized in Supplementary Table 1. All structural figures were generated using Pymol (http://www.pymol.org).

Secondary-structure determination. The secondary structure determination was based on the ESPript[38] algorithm.

29. Zhang, W. *et al.* Crystal structure of the swine-origin A (H1N1)-2009 influenza A virus hemagglutinin (HA) reveals similar antigenicity to that of the 1918 pandemic virus. *Protein Cell* **1,** 459–467 (2010).
30. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
31. Collaborative Computing Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
32. Usón, I. & Sheldrick, G. M. Advances in direct methods for protein crystallography. *Curr. Opin. Struct. Biol.* **9,** 643–648 (1999).
33. Read, R. J. Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr. D* **57,** 1373–1382 (2001).
34. Cowtan, K. D. & Zhang, K. Y. Density modification for macromolecular phase improvement. *Prog. Biophys. Mol. Biol.* **72,** 245–270 (1999).
35. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66,** 213–221 (2010).
36. Emsley, P. & Cowtan, K. *Coot*: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
37. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. *PROCHECK*: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26,** 283–291 (1993).
38. Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15,** 305–308 (1999).

LETTER

T_{reg} induction by a rationally selected mixture of Clostridia strains from the human microbiota

Koji Atarashi[1,2,3]*, Takeshi Tanoue[1,2]*, Kenshiro Oshima[4,5]*, Wataru Suda[5], Yuji Nagano[1,2], Hiroyoshi Nishikawa[6], Shinji Fukuda[1,7], Takuro Saito[6], Seiko Narushima[1], Koji Hase[1,3], Sangwan Kim[5], Joëlle V. Fritz[8], Paul Wilmes[8], Satoshi Ueha[9], Kouji Matsushima[9], Hiroshi Ohno[1], Bernat Olle[10], Shimon Sakaguchi[6], Tadatsugu Taniguchi[2], Hidetoshi Morita[4,11], Masahira Hattori[5] & Kenya Honda[1,2,4]

Manipulation of the gut microbiota holds great promise for the treatment of inflammatory and allergic diseases[1,2]. Although numerous probiotic microorganisms have been identified[3], there remains a compelling need to discover organisms that elicit more robust therapeutic responses, are compatible with the host, and can affect a specific arm of the host immune system in a well-controlled, physiological manner. Here we use a rational approach to isolate CD4+FOXP3+ regulatory T (T_{reg})-cell-inducing bacterial strains from the human indigenous microbiota. Starting with a healthy human faecal sample, a sequence of selection steps was applied to obtain mice colonized with human microbiota enriched in T_{reg}-cell-inducing species. From these mice, we isolated and selected 17 strains of bacteria on the basis of their high potency in enhancing T_{reg} cell abundance and inducing important anti-inflammatory molecules—including interleukin-10 (IL-10) and inducible T-cell co-stimulator (ICOS)—in T_{reg} cells upon inoculation into germ-free mice. Genome sequencing revealed that the 17 strains fall within clusters IV, XIVa and XVIII of Clostridia, which lack prominent toxins and virulence factors. The 17 strains act as a community to provide bacterial antigens and a TGF-β-rich environment to help expansion and differentiation of T_{reg} cells. Oral administration of the combination of 17 strains to adult mice attenuated disease in models of colitis and allergic diarrhoea. Use of the isolated strains may allow for tailored therapeutic manipulation of human immune disorders.

CD4+FOXP3+ T_{reg} cells are present most abundantly in the intestinal mucosa at steady state, and contribute to intestinal and systemic immune homeostasis[4–7]. In germ-free mice, the frequency of colonic T_{reg} cells and levels of IL-10 expression by T_{reg} cells are markedly reduced[4–7]. We have shown previously that a combination of Clostridia strains isolated from conventionally reared mice potently affect the number and function of CD4+FOXP3+ T_{reg} cells in mouse colonic lamina propria[4]. In an attempt to enable clinical translation of our previous findings, we aimed to identify T_{reg}-cell-inducing bacterial strains derived from the human microbiota (see Supplementary Fig. 1 for a summary of the procedure).

We obtained a human stool sample from a healthy Japanese volunteer. Because we previously reported that the chloroform-resistant fraction of mouse gut microbiota was enriched in T_{reg}-cell-inducing species[4], the stool sample was either untreated or treated with chloroform and orally inoculated into IQI/Jic germ-free mice. Each group of ex-germ-free (exGF) mice was separately housed for 3–4 weeks in vinyl isolators to avoid further microbial contamination. Although a recent study showed that the human microbiota had no impact on the

immune responses in the mouse small intestine[8], we observed a significant increase in the percentage of FOXP3+ T_{reg} cells among CD4+ T cells in the colons of exGF mice inoculated with untreated human faeces compared with germ-free mice (Fig. 1a and Supplementary Fig. 2). Notably, a more pronounced increase was observed in the colons of exGF mice inoculated with chloroform-treated human faeces (Fig. 1a). These findings suggest that the human intestinal microbiota contains T_{reg}-cell-inducing bacteria, and that they are enriched in the chloroform-resistant fraction. We also examined the effects of human faeces inoculation on colonic IL-17- and IFN-γ-expressing CD4+ cells (T_H17 and T_H1 cells). In exGF mice inoculated with untreated or chloroform-treated human faeces, the frequency of T_H1 cells was unchanged compared with germ-free mice (Fig. 1b). By contrast, there
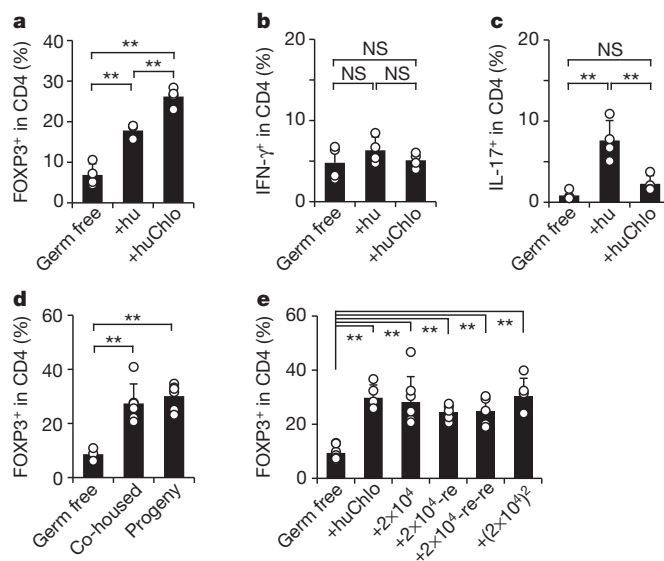

Figure 1 | T_{reg} cell accumulation in germ-free mice induced by inoculation with human microbiota. a–e, The percentages of FOXP3+, IL-17+ and IFN-γ+ cells within the CD4+ cell population in the colon lamina propria of the indicated mice are shown (see also Supplementary Fig. 2). Circles represent individual animals. The height of the black bars indicates the mean. All experiments were performed more than twice with similar results. Error bars indicate s.d. **P < 0.01; NS, not significant. +hu, exGF mice inoculated with untreated human faeces; +huChlo, exGF mice inoculated with chloroform-treated human faeces. (See the main text for further definitions of x-axis labels.)

[1]RIKEN Center for Integrative Medical Sciences (IMS-RCAI), 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan. [2]Department of Immunology, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. [3]PRESTO, Japan Science and Technology Agency, 4-1-8 Honcho Kawaguchi, Saitama 332-0012, Japan. [4]CREST, Japan Science and Technology Agency, 4-1-8 Honcho Kawaguchi, Saitama 332-0012, Japan. [5]Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8561, Japan. [6]Experimental Immunology, Immunology Frontier Research Center, Osaka University, 3-1 Yamadaoka, Suita, Osaka 565-0871, Japan. [7]Institute for Advanced Biosciences, Keio University, 246-2 Mizukami, Tsuruoka, Yamagata 997-0052, Japan. [8]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Avenue des Hauts-Fourneaux, 7, Esch-sur-Alzette, L-4362, Luxembourg. [9]Department of Molecular Preventive Medicine, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. [10]PureTech Ventures, 500 Boylston Street, Suite 1600, Boston, Massachusetts 02116, USA. [11]School of Veterinary Medicine, Azabu University, 1-17-71 Fuchinobe, Sagamihara, Kanagawa 252-5201, Japan.
*These authors contributed equally to this work.

was a significant accumulation of T_H17 cells in the colons of exGF mice inoculated with untreated human faeces (Fig. 1c and Supplementary Fig. 2). Notably, the capacity of human faeces to induce T_H17 cells was greatly diminished after treatment with chloroform (Fig. 1c). These results indicate that the chloroform-sensitive bacterial fraction in the human stool tested contained T_H17-cell-inducing bacteria, whereas the chloroform-resistant bacteria preferentially promoted T_{reg} cell accumulation in the colon.

To investigate whether T_{reg} cell induction by the chloroform-resistant fraction of human intestinal bacteria is transmissible, adult germ-free mice were co-housed with exGF mice inoculated with chloroform-treated human faeces for 4 weeks. Co-housed mice showed a significant increase in the frequency of colonic T_{reg} cells (Fig. 1d). In addition, the progeny of exGF mice inoculated with chloroform-treated human faeces also showed increased numbers of T_{reg} cells (Fig. 1d). Therefore, T_{reg} cell induction by human intestinal bacteria is horizontally and vertically transmissible. Oral inoculation of germ-free mice with 2×10^4-fold diluted caecal samples from exGF mice inoculated with chloroform-treated human faeces fully induced the accumulation of T_{reg} cells in the colon lamina propria, suggesting that abundant rather than minor members of the intestinal microbiota in exGF mice inoculated with chloroform-treated human faeces drive the observed induction of T_{reg} cells (Fig. 1e). The T_{reg}-cell-inducing microbiota in mice inoculated with the 2×10^4-fold diluted sample ($+2 \times 10^4$ mice) was a stable community, because serial oral inoculation of caecal contents

from these mice equally induced the accumulation of T_{reg} cells in secondary ($+2 \times 10^4$-re mice) and tertiary recipients ($+2 \times 10^4$-re-re mice) (Fig. 1e). To minimize nonessential components of the microbiota for T_{reg} cell induction, the caecal contents of $+2 \times 10^4$ mice were again diluted 2×10^4-fold and orally inoculated into another set of germ-free mice ($+(2 \times 10^4)^2$ mice). The $+(2 \times 10^4)^2$ mice had a marked accumulation of T_{reg} cells in the colon (Fig. 1e). These results suggested that we succeeded in obtaining mice colonized with a relatively restricted and stable community of bacterial species enriched for T_{reg} cell inducers.

The composition of the gut microbiota in mice treated with human samples was analysed by 16S ribosomal RNA (rRNA) gene amplicon sequencing using a 454 sequencer. Quality filter-passed sequences (3,000 reads for each sample) were classified into operational taxonomic units (OTUs) based on sequence similarity (>96% identity). The numbers of detected reads and closest known species for each OTU are shown in Supplementary Table 1, and the relative abundance of OTUs in each caecal sample is shown in Fig. 2a. As expected, the OTU profiles of mice treated with human faeces were quite different from those of conventional specific pathogen-free (SPF) mice (Supplementary Fig. 3). In mice inoculated with untreated human faeces, OTUs belonging to Bacteroidetes accounted for about 50% of the caecal microbial community (Fig. 2a). By contrast, most OTUs in exGF mice inoculated with chloroform-treated human faeces were related to Clostridia species. Most bacteria in $+2 \times 10^4$, $+2 \times 10^4$-re and $+(2 \times 10^4)^2$ mice had 16S rRNA gene sequence similarities with about 20 species of Clostridia, listed in Fig. 2b.

To isolate bacterial strains with T_{reg}-cell-inducing capabilities, we cultured caecal contents from $+2 \times 10^4$, $+2 \times 10^4$-re and $+(2 \times 10^4)^2$ mice in vitro and picked 442 colonies. BLAST searches of 16S rRNA gene sequences of the isolated colonies revealed that 31 strains in total were present, all of which were Clostridia (Supplementary Fig. 4). Of the 31 strains, we selected 23 that had less than 99% 16S rRNA gene sequence identity to any of the other 30 strains (Supplementary Fig. 4). We then individually cultured the 23 strains, mixed them in equal amounts, and orally inoculated the mixture into germ-free IQI mice ($+23$-mix mice). Numerous rod- and round-shaped bacteria were observed by scanning electron microscopy (SEM) on the epithelial cell surface in $+23$-mix mice (Fig. 2c), and the size and appearance of the caeca were quite different from those in germ-free mice, indicating successful colonization (Supplementary Fig. 5a). Pyrosequencing of 16S rRNA genes revealed that the caecal microbiota composition in $+23$-mix mice was quite similar to that in $+(2 \times 10^4)^2$ mice (Fig. 2a). In $+23$-mix mice, we observed efficient induction of T_{reg} cells in the colonic lamina propria (Fig. 2d). The magnitude was comparable to that observed in exGF mice inoculated with chloroform-treated human faeces and much higher than that in mice colonized with *Faecalibacterium prausnitzii*, a human Clostridia strain well known for enhancing regulatory cell functions[9] (Fig. 2d). Most T_{reg} cells in $+23$-mix mice expressed low levels of Helios (also known as IKZF2), indicating antigen-experienced cells (Fig. 2e, Supplementary Fig. 5b and ref. 10).

Only 17 strains listed in Fig. 2b and Supplementary Fig. 4 were detected in $+23$-mix mice by 16S rRNA gene sequencing, indicating that these 17 strains may be sufficient to induce T_{reg} cells. Indeed, we found that the mixture of 17 strains (17-mix) induced FOXP3$^+$ T_{reg} cells to a similar extent as the 23-mix (Fig. 3a). The increase in T_{reg} cells induced by the 17-mix was reproducibly observed in exGF mice of different genetic backgrounds (IQI, BALB/c and C57BL/6) (Fig. 3a). Moreover, the mix was effective in other rodents: the frequency of colonic T_{reg} cells in exGF rats inoculated with 17-mix was significantly higher than that in germ-free rats and comparable to that in SPF rats (Fig. 3a). The colonization with 17-mix induced a significant increase in the frequency of IL-10$^+$ and/or ICOS$^+$ cells within the T_{reg} cell population, as revealed by analysis of exGF IL-10 reporter mice (*Il10*^{Venus} mice, ref. 4) colonized with the 17-mix (Fig. 3b). Furthermore, IL-10$^+$ T_{reg} cells in
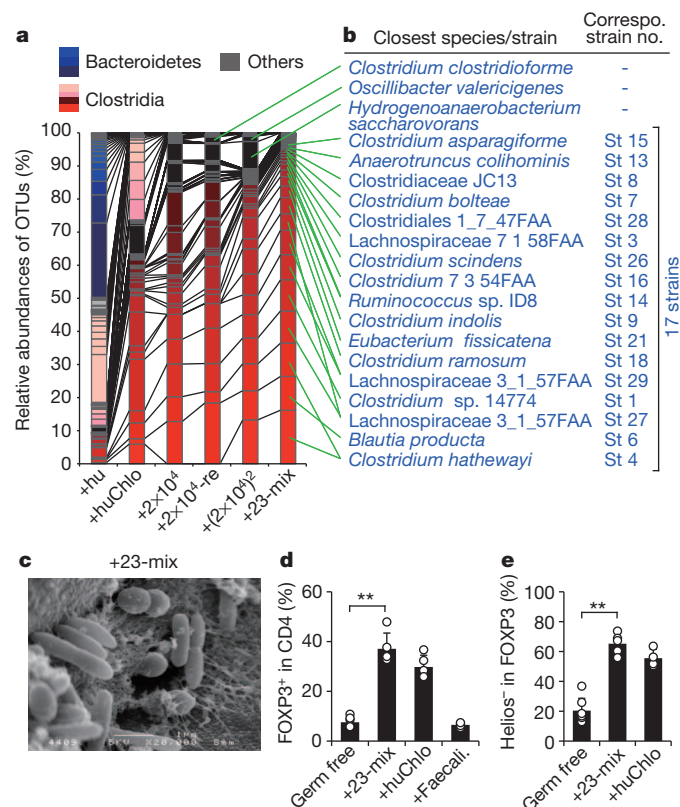

Figure 2 | Assessment of microbiota composition and isolation of T_{reg}-cell-inducing strains. a, b, Pyrosequencing of 16S rRNA genes was performed on caecal contents from the indicated mice. Relative abundance of OTUs (%) in the caecal bacterial community in each mouse (**a**), and the closest species/strain in the database and the corresponding isolated strain number for the indicated OTU (**b**) are shown. **c,** SEM showing the proximal colon of $+23$-mix mice. Original magnification, $\sim \times 20{,}000$. **d, e,** The percentages of FOXP3$^+$ cells within the CD4$^+$ cell population (**d**) and Helios$^-$ cells in CD4$^+$FOXP3$^+$ cells (**e**) in the colon of the indicated mice. Circles represent individual animals. All experiments were performed more than twice with similar results. Error bars indicate s.d. **P < 0.01.
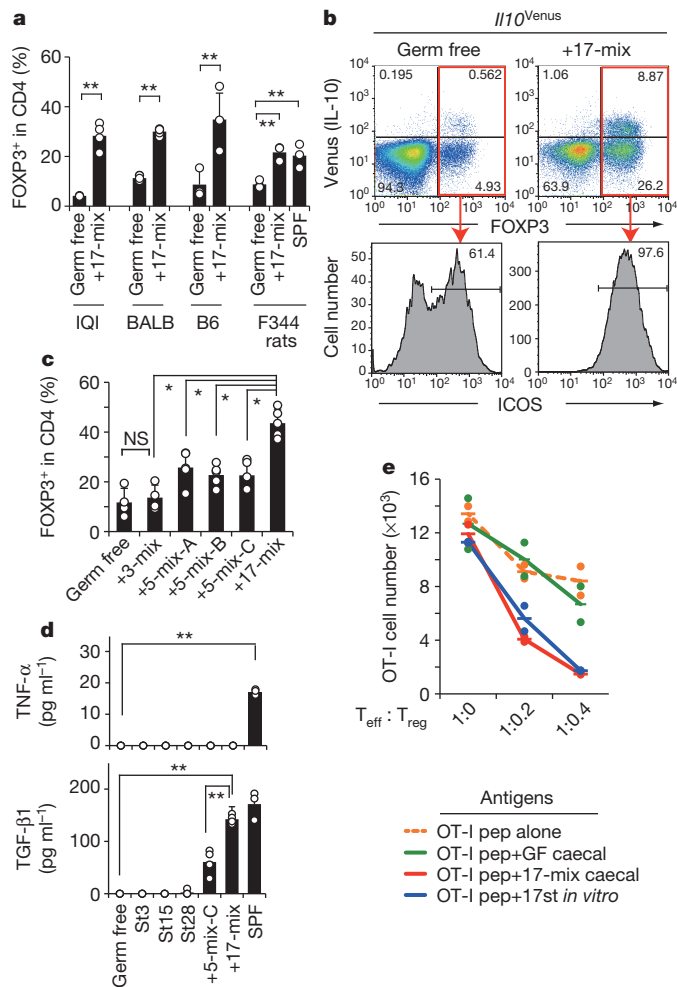
Figure 3 | Characterization of 17 T_{reg}-cell-inducing strains. a, The percentages of FOXP3$^+$ cells within the CD4$^+$ cell population in the colon lamina propria of the indicated mice and rats. **b**, The expression of Venus (IL-10) and FOXP3 by the gated colonic lamina propria CD4$^+$ cells, and ICOS expression by CD4$^+$FOXP3$^+$ cells in exGF $Il10^{Venus}$ mice colonized with or without 17-mix. **c**, Percentages of FOXP3$^+$ cells within the CD4$^+$ cell population in IQI exGF mice colonized with the indicated mix. **d**, The production of TNF-α and TGF-β1 in HCT8 cells stimulated with caecal extracts from the indicated mice. **e**, CD8$^+$ T cells from OT-I mice (T_{eff}) and the indicated ratio of colon lamina propria CD4$^+$CD25$^+$ cells from +17-mix mice (T_{reg}) were incubated with CD11c$^+$ cells pulsed with OT-I peptide alone or in combination with autoclaved caecal contents from +17-mix mice (+17-mix caecal), germ-free mice (+GF caecal), or autoclaved 17 strains cultured in vitro (+17 st in vitro). Depicted data represent average of duplicates (see also Supplementary Fig. 9c). Circles in **a**, **c**–**e** represent samples from individual animals. All experiments were performed more than twice with similar results. Error bars indicate s.d. **$P < 0.01$; *$P < 0.05$; NS, not significant.

+17-mix mice expressed high levels of CTLA4 (Supplementary Fig. 5c). Because IL-10 and CTLA4 are essential for the immunosuppressive activity of T_{reg} cells[11,12], and ICOS is required for the T_{reg}-cell-mediated suppression of T_H2 responses[13], these results suggest that the mixture of 17 strains affects both the number and function of T_{reg} cells in the colon. Next, we monocolonized germ-free mice with one of each of the 17 individual strains to determine their individual T_{reg} cell induction capability. The monocolonized exGF mice exhibited low to intermediate levels of T_{reg} cell induction with inter-individual variability (Supplementary Fig. 6a). As expected, none of the strains induced T_H17 cells in the monocolonized mice (Supplementary Fig. 6b). We also examined T_{reg} cell induction by subsets of the 17-mix (randomly selected combinations of 3–5 strains: 3-mix, 5-mix-A, 5-mix-B, and 5-mix-C, see Supplementary Fig. 4). Although all tested combinations of 5-mix induced increases

in the frequency of T_{reg} cells, the magnitude was substantially lower than that observed in +17-mix mice (Fig. 3c). Therefore, it is likely that the 17 strains act synergistically to amplify the induction of T_{reg} cells in a microbial-community-dependent fashion.

To investigate the mechanism for the T_{reg} cell induction by the community of 17 strains, we incubated various human and mouse intestinal epithelial cell lines and primary cells with aqueous extracts from caecal contents from the +17-mix mice, and assessed the production of the active form of TGF-β1, a key cytokine for the differentiation and expansion of T_{reg} cells. The caecal extracts from +17-mix mice routinely elicited TGF-β1, but not IL-6 and TNF-α, production, and the magnitude was significantly higher than that elicited by caecal extracts from single-strain or 5-mix-colonized mice (Fig. 3d and Supplementary Fig. 7). The induction of TGF-β1 was not inhibited by pre-treatment of the caecal extracts with a protease or nuclease (Supplementary Fig. 7c). Short-chain fatty acids (SCFAs) are protease- and nuclease-insensitive and have been associated with regulation of host immune homeostasis[14]. Quantitative analysis of SCFAs in caecal contents from +17-mix mice showed a significantly higher concentration of acetate, propionate, butyrate and isobutyrate than that in single-strain or 5-mix-colonized mice (Supplementary Fig. 8a). Furthermore, a mixture of sodium salts of these SCFAs elicited TGF-β1 production in epithelial cells to a level similar to that seen when the cells were stimulated with caecal extracts from +17-mix mice (Supplementary Fig. 8b). Therefore, the community of 17 strains cooperatively produces SCFAs that can elicit a TGF-β response, and this activity may contribute to the differentiation and expansion of T_{reg} cells. We also investigated whether the 17 strains provide bacterial antigens to T cells. To do this, we addressed the antigen specificity of T_{reg} cells accumulated in +17-mix mice using a cognate antigen-driven suppression assay. CD4$^+$CD25$^+$ lamina propria T cells from +17-mix mice substantially inhibited the OT-I ovalbumin (OVA) peptide-driven proliferation of OT-I CD8 T cells, and this suppression was markedly enhanced in the presence of autoclaved caecal content from +17-mix mice or autoclaved 17 strains cultured in vitro, but not in the presence of OT-II OVA peptide or caecal content from germ-free mice (Fig. 3e and Supplementary Fig. 9). These results are consistent with previous reports[15,16] and suggest that some fraction of colonic lamina propria T_{reg} cells in +17-mix mice is specific to the 17 strains of Clostridia. Next, we assessed the kinetics of T_{reg} cell accumulation and their expression of Ki67, a cell-cycle-associated nuclear protein, and gut-homing-associated molecules CD103 and β7 integrin. We observed a marked increase in the proportion of Ki67, CD103 and β7 expressing cells by 1 week after inoculation with the 17-mix (Supplementary Figs 10 and 11). Collectively, these observations indicate that the 17 strains provide SCFAs, bacterial antigens and probably other factors, which together contribute to differentiation, expansion and colonic homing of T_{reg} cells.

To define the identity of the 17 bacterial strains fully, we sequenced their genomes (Supplementary Fig. 12). Phylogenetic comparison of the 17 strains using ribosomal multi locus sequencing typing (rMLST) revealed that the 17 strains belong to bacterial species falling within clusters XIVa, IV and XVIII of Clostridia as defined previously[17] (in a recent taxonomy, members of cluster XVIII Clostridia were reclassified in the class Erysipelotrichi) (Supplementary Fig. 13). The genome sequencing also revealed that the 17 strains all lack strong virulence-related genes such as collagenase and phospholipase C, often identified in pathogenic Clostridia species (Supplementary Table 2). We then examined the relative abundance of the 17 strains in healthy and ulcerative colitis human subjects using draft genome sequences of the 17 strains and publicly available human microbiome genomes generated through the MetaHIT project[18]. Ulcerative colitis subjects showed a tendency towards a reduction of the 17 strains, and 5 out of the 17 strains were significantly reduced in ulcerative colitis patients (Supplementary Fig. 14).

To evaluate the potential benefits of supplementation with the 17 strains, 17-mix or control PBS was orally administered into adult SPF

mice every 2 or 3 days (SPF + 17-mix or SPF + ctrl, respectively). We confirmed a significant increase in the frequency of colonic T_{reg} cells in SPF + 17-mix mice compared with SPF + ctrl mice after 3 weeks of treatment (Fig. 4a). While being continuously treated with 17-mix or control, mice were subjected to the OVA-induced allergic diarrhoea model[19]. The occurrence and severity of diarrhoea and the OVA-specific serum IgE levels were significantly reduced in SPF + 17-mix mice relative to control mice (Fig. 4b–d). The protective effect of 17-mix was significantly attenuated by treatment of mice with a T_{reg}-cell-depleting anti-CD25 antibody (Supplementary Fig. 15). We also subjected mice to an experimental colitis model induced by trinitrobenzene sulphonic acid (TNBS)[20]. SPF + 17-mix mice showed less severe colon shortening and milder histological disease features, accompanied by lower mortality

than control mice (Fig. 4e–g and Supplementary Fig. 16a). In keeping with this clinical outcome, there was significantly increased expression of *Foxp3* and *Tgfb1* mRNA in SPF + 17-mix mice compared with control mice, as well as a tendency towards a reduction of inflammatory cytokine transcripts (Supplementary Fig. 16b). Identical suppression of colitis by 17-mix was also observed in an adoptive transfer model, in which germ-free SCID mice were orally inoculated with faeces from SPF mice together with or without 17-mix and then transferred with $CD4^+CD45RB^{hi}$ T cells (Supplementary Fig. 17).

The clinical track record of efficacy of single-strain probiotics has been modest. It has been postulated that a collection of functionally distinct bacterial species rationally selected from the human gut microbiota may be more effective than single strains in preventing/treating disease[21]. In the present study, we isolated 17 strains within Clostridia clusters XIVa, IV and XVIII from a human faecal sample; these strains affect T_{reg} cell differentiation, accumulation and function in the mouse colon. It remains to be seen whether the 17 strains will have similar effects in the human intestine; however, a decreased prevalence of Clostridia clusters XIVa and IV in faecal samples from patients with inflammatory bowel disease and atopy[22–24] may suggest that supplementation with the 17-strain bacterial community might counterbalance dysbiosis, induce T_{reg} cells and aid in the management of allergic and inflammatory conditions.
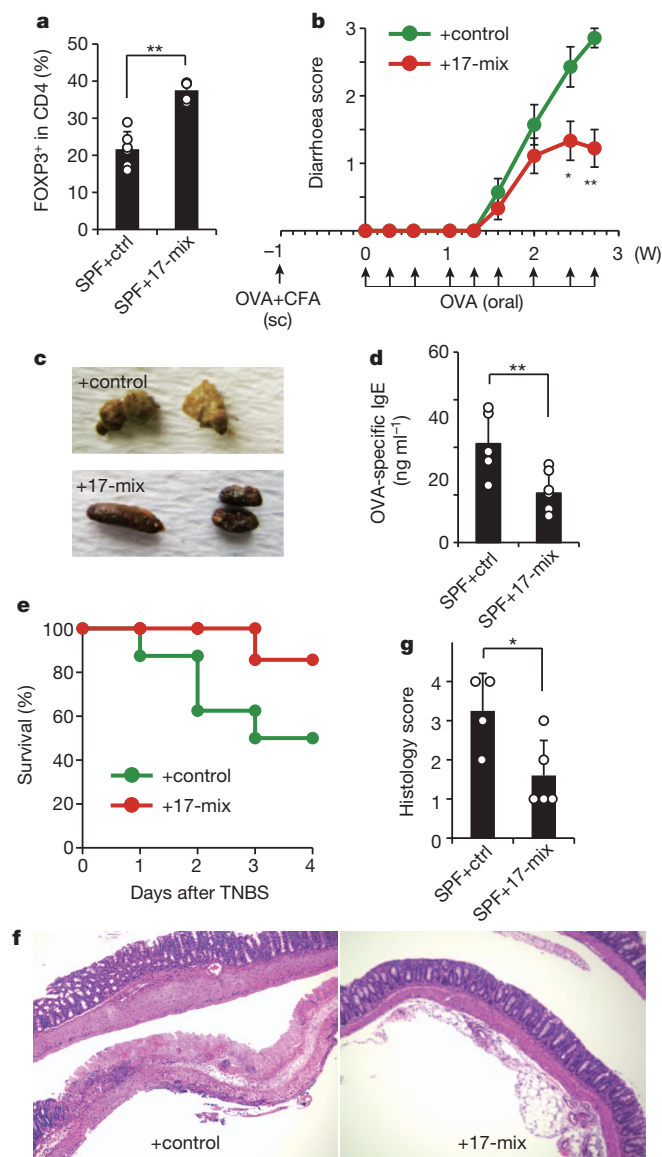
METHODS SUMMARY

Figure 4 | Treatment with 17-mix suppresses experimental colitis models. a, The percentages of $FOXP3^+$ cells within the $CD4^+$ cell population in SPF + 17-mix or SPF + ctrl mice. **b–d,** SPF + 17-mix ($n = 9$) and SPF + ctrl ($n = 7$) mice were subjected to OVA-induced diarrhoea. The diarrhoea score (**b**; see Methods for definition), representative photographs of faeces (**c**), and OVA-specific IgE levels in the sera (**d**) are shown. sc, subcutaneous. **e–g,** SPF + 17-mix ($n = 8$) and SPF + ctrl ($n = 7$) were treated with TNBS. Animal survival (**e**), haematoxylin and eosin staining (original magnification, ×10) (**f**), and histology score of the distal colon (**g**) on day 4 after TNBS administration are shown. Data are representative of two independent experiments. Error bars indicate s.d. **$P < 0.01$; *$P < 0.05$.

1. Round, J. L. & Mazmanian, S. K. The gut microbiota shapes intestinal immune responses during health and disease. *Nature Rev. Immunol.* **9,** 313–323 (2009).
2. Honda, K. & Littman, D. R. The microbiome in infectious disease and inflammation. *Annu. Rev. Immunol.* **30,** 759–795 (2012).
3. O'Toole, P. W. & Cooney, J. C. Probiotic bacteria influence the composition and function of the intestinal microbiota. *Interdiscip. Perspect. Infect. Dis.* **2008,** 175285 (2008).
4. Atarashi, K. *et al.* Induction of colonic regulatory T cells by indigenous *Clostridium* species. *Science* **331,** 337–341 (2011).
5. Geuking, M. B. *et al.* Intestinal bacterial colonization induces mutualistic regulatory T cell responses. *Immunity* **34,** 794–806 (2011).
6. Russell, S. L. *et al.* Early life antibiotic-driven changes in microbiota enhance susceptibility to allergic asthma. *EMBO Rep.* **13,** 440–447 (2012).
7. Round, J. L. & Mazmanian, S. K. Inducible Foxp3$^+$ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. *Proc. Natl Acad. Sci. USA* **107,** 12204–12209 (2010).
8. Chung, H. *et al.* Gut immune maturation depends on colonization with a host-specific microbiota. *Cell* **149,** 1578–1593 (2012).
9. Sokol, H. *et al.* *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl Acad. Sci. USA* **105,** 16731–16736 (2008).
10. Thornton, A. M. *et al.* Expression of Helios, an Ikaros transcription factor family member, differentiates thymic-derived from peripherally induced Foxp3$^+$ T regulatory cells. *J. Immunol.* **184,** 3433–3441 (2010).
11. Rubtsov, Y. P. *et al.* Regulatory T cell-derived interleukin-10 limits inflammation at environmental interfaces. *Immunity* **28,** 546–558 (2008).
12. Wing, K. *et al.* CTLA-4 control over Foxp3$^+$ regulatory T cell function. *Science* **322,** 271–275 (2008).
13. Zheng, Y. *et al.* Regulatory T-cell suppressor program co-opts transcription factor IRF4 to control T_H2 responses. *Nature* **458,** 351–356 (2009).
14. Maslowski, K. M. & Mackay, C. R. Diet, gut microbiota and immune responses. *Nature Immunol.* **12,** 5–9 (2011).
15. Lathrop, S. K. *et al.* Peripheral education of the immune system by colonic commensal microbiota. *Nature* **478,** 250–254 (2011).
16. Cebula, A. *et al.* Thymus-derived regulatory T cells contribute to tolerance to commensal microbiota. *Nature* **497,** 258–262 (2013).

17. Collins, M. D. *et al.* The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations. *Int. J. Syst. Bacteriol.* **44,** 812–826 (1994).
18. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464,** 59–65 (2010).
19. Kweon, M. N., Yamamoto, M., Kajiki, M., Takahashi, I. & Kiyono, H. Systemically derived large intestinal CD4$^+$ Th2 cells play a central role in STAT6-mediated allergic diarrhea. *J. Clin. Invest.* **106,** 199–206 (2000).
20. Strober, W., Fuss, I. J. & Blumberg, R. S. The immunology of mucosal models of inflammation. *Annu. Rev. Immunol.* **20,** 495–549 (2002).
21. Lawley, T. D. *et al.* Targeted restoration of the intestinal microbiota with a simple, defined bacteriotherapy resolves relapsing *Clostridium difficile* disease in mice. *PLoS Pathog.* **8,** e1002995 (2012).
22. Frank, D. N. *et al.* Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc. Natl Acad. Sci. USA* **104,** 13780–13785 (2007).
23. Manichanh, C. *et al.* Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* **55,** 205–211 (2006).
24. Candela, M. *et al.* Unbalance of intestinal microbiota in atopic children. *BMC Microbiol.* **12,** 95 (2012).

Supplementary Information is available in the online version of the paper.

Author Contributions K.Ho. planned experiments, analysed data and wrote the paper together with B.O. and M.H.; K.A. and T.Tano. performed immunological analyses and bacterial cultures together with Y.N., S.N. and H.M.; W.S., K.O., S.K. and M.H. performed bacterial sequence analyses; K.M. and S.U. provided essential materials; H.N., T.S. and S.S. supervised the T$_{reg}$ cell suppression assay; S.F., K.Ha., H.O., T.Tani., J.V.F. and P.W. were involved in data discussions.

Author Information All genome sequence data are deposited in DDBJ BioProject ID PRJDB521-543. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.H. (hattori@k.u-tokyo.ac.jp) or K.Ho. (kenya@rcai.riken.jp).

METHODS

Mice and rats. C57BL/6, BALB/c, IQI/Jic and CB.17 SCID mice and F344 rats kept under SPF or germ-free conditions were purchased from Sankyo laboratories, Japan SLC, or CLEA Japan. IQI germ-free mice were used unless otherwise indicated. Germ-free and gnotobiotic mice were bred and maintained in vinyl isolators within the gnotobiotic facility of Sankyo laboratories. Germ-free $Il10^{Venus}$ mice were generated as previously described[4]. OT-I and OT-II T-cell receptor transgenic mice were purchased from Taconic Farms. All animal experiments were approved by the Animal Research Committee of RIKEN Yokohama Institute and the University of Tokyo.

Chloroform treatment of human stool and generation of gnotobiotic mice. Human stool from a healthy volunteer (Japanese, male, age 31 years) was obtained with informed consent. Human stool and mouse caecal contents were directly frozen at $-80\,^{\circ}C$, or suspended in 4 times volume (w/v) of phosphate-buffered saline (PBS) + 20% glycerol solution, snap-frozen in liquid nitrogen and stored at $-80\,^{\circ}C$ until use. The frozen stocks were thawed, suspended in 10 times volume (w/v) of PBS, and passed through a 70 µm cell strainer to eliminate clumps and debris. Then suspensions were mixed with chloroform (final concentration 3%), and incubated in a shaking water bath for 60 min. After evaporation of chloroform by bubbling with N_2 gas for 30 min, the aliquots containing the chloroform-resistant fraction of intestinal bacteria were inoculated into germ-free mice by intra-gastric administration (250 µl; per mouse). To generate a series of gnotobiotic mice inoculated with diluted samples, caecal contents from exGF mice were treated with chloroform, diluted with PBS, and inoculated into germ-free IQI mice. The caecal suspensions diluted 2×10^4-fold correspond to 2.5×10^4 bacterial cells per mouse. Each group of exGF mice was individually caged in the gnotobiotic isolator for 3–4 weeks at Sankyo Lab service.

Isolation of intestinal lamina propria lymphocytes and flow cytometry. The colons were collected and opened longitudinally, washed with PBS to remove all luminal contents and shaken in Hanks' balanced salt solution (HBSS) containing 5 mM EDTA for 20 min at $37\,^{\circ}C$. After removing epithelial cells, muscle layers and fat tissue using forceps, the lamina propria layers were cut into small pieces and incubated with RPMI1640 containing 4% fetal bovine serum, $0.5\,mg\,ml^{-1}$ collagenase D, $0.5\,mg\,ml^{-1}$ dispase and $40\,\mu g\,ml^{-1}$ DNase I (all Roche Diagnostics) for 1 h at $37\,^{\circ}C$ in a shaking water bath. The digested tissues were washed with HBSS containing 5 mM EDTA, resuspended in 5 ml of 40% Percoll (GE Healthcare) and overlaid on 2.5 ml of 80% Percoll in a 15-ml Falcon tube. Percoll gradient separation was performed by centrifugation at $800g$ for 20 min at $25\,^{\circ}C$. The lamina propria lymphocytes were collected from the interface of the Percoll gradient and suspended in ice-cold PBS. For analysis of T_{reg} cells, isolated lymphocytes were labelled with the LIVE/DEAD fixable dead cell stain kit (Invitrogen) to exclude dead cells from the analysis. The cells were washed with staining buffer containing PBS, 2% FBS, 2 mM EDTA and 0.09% NaN_3 and surface staining was performed with PECy7- or Pacific blue-labelled anti-CD4 antibody (RM4-5, BD Biosciences), PE-labelled anti-ICOS antibody (C938.4A, BioLegend), Alexa488-labelled anti-CD103 antibody (2E7, BioLegend), and PerCP/Cy5.5-labelled anti-integrin-β7 antibody (FIB27, BioLegend). Intracellular staining of FOXP3, CTLA4, Helios and Ki67 was performed using the Alexa647-labelled anti-FOXP3 antibody (FJK-16 s, eBioscience), PE-labelled anti-CTLA4 antibody (UC10-4F10-11, BD Biosciences), PE-labelled anti-Helios antibody (22F6, BioLegend), PECy7-labelled anti-Ki67 antibody (B56, BD Biosciences) and FOXP3 staining buffer set (eBioscience). For analysis of T_H1 and T_H17 cells, isolated lymphocytes were stimulated for 4 h with $50\,ng\,ml^{-1}$ phorbol 12-myristate 13-acetate (PMA, Sigma) and $1\,\mu g\,ml^{-1}$ ionomycin (Sigma) in the presence of GolgiStop (BD Biosciences). After incubation for 4 h, cells were washed in PBS, labelled with the LIVE/DEAD fixable dead cell stain kit and surface CD4 was stained with PECy7-labelled anti-CD4 antibody. Cells were washed, fixed in Cytofix/Cytoperm, permeabilized with Perm/Wash buffer (BD Biosciences), and stained with the APC-labelled anti-IL-17 antibody (eBio17B7, eBioscience) and FITC-labelled anti-IFN-γ antibody (XMG1.2, BD Biosciences). The antibody-stained cells were analysed with LSR Fortessa or FACSAriaIII (BD Biosciences), and data were analysed using FlowJo software (Treestar).

Meta 16S rRNA gene sequencing. The caecal contents from exGF mice were suspended in 10 ml of Tris-EDTA containing 10 mM Tris-HCl and 1 mM EDTA (pH 8), and incubated with lysozyme (Sigma, $15\,mg\,ml^{-1}$) at $37\,^{\circ}C$ for 1 h with gentle mixing. A purified achromopeptidase (Wako) was added (final concentration 2,000 unit ml^{-1}) and further incubated at $37\,^{\circ}C$ for another 30 min. Then, sodium dodecyl sulphate (final concentration 1%) was added to the cell suspension and mixed well. Subsequently, proteinase K (Merck) was added (final concentration $1\,mg\,ml^{-1}$) to the suspension and the mixture was incubated at $55\,^{\circ}C$ for 1 h. High-molecular-mass DNA was isolated and purified by phenol/chloroform extraction, ethanol and finally polyethyleneglycol precipitation[25]. PCR was performed using Ex Taq (TAKARA) and (1) the 454 primer A (5′-CCATCTCA

TCCCTGCGTGTCTCCGACTCAG (454 adaptor sequence) + barcode (10 bases) + AGRGTTTGATYMTGGCTCAG-3′ (27Fmod)) and (2) the 454 primer B (5′-CCTATCCCCTGTGTGCCTTGGCAGTCTCAG (454 adaptor sequence) + TGCTGCCTCCCGTAGGAGT-3′ (338R)) to the V1–V2 region of the 16S rRNA gene. Amplicons generated from each sample (~330 bp) were subsequently purified using AMPur XP (Beckman Coulter). The amount of DNA was quantified using Quant-iT Picogreen dsDNA assay kit (Invitrogen) and TBS-380mini fluorometer (Turner Biosystems). Then, the amplified DNA was used as template for 454 GS Junior (Roche) pyrosequencing using GS Junior Titanium emPCR Kit-Lib-L, GS Junior Titanium Sequencing Kit and GS Junior Titanium PicoTiterPlate Kit (all Roche) according to the manufacturer's instructions. Quality-filter-passed reads were obtained by removing reads that did not have both primer sequences, had the average quality value (QV) <25, and were possibly chimaeric[26]. Of the filter-passed reads, 3,000 reads trimming off both primer sequences for each sample were used and subjected to OTU analysis with the cutoff similarity of 96% identity. Representative sequences from each OTU were blasted to Ribosomal Database Project (RDP) of bacterial isolates, our genome database constructed from publicly available genome sequences in NCBI and HMP databases, and 16S sequences of the 23 strains obtained in this study.

Isolation of bacterial strains. The frozen stocks of caecal contents from exGF mice were serially diluted with PBS and seeded onto non-selective agar plates (blood liver (BL) agar (Eiken Chemical) or Eggerth-Gagnon (EG) agar plates). EG agar plates contain the following components (quantities expressed per litre): Lab-Lemco Powder (2.8 g, Oxoid); proteose peptone no. 3 (10.0 g, Difco); yeast extract (5.0 g, Difco); Na_2HPO_4 (4.0 g); D(+)-glucose (1.5 g); soluble starch (0.5 g); L-cystine (0.2 g); L-cysteine-HCl-H_2O (0.5 g); Tween 80 (0.5 g); Bacto agar (16.0 g, Difco); and defibrinated horse blood (50 ml). After culture under aerobic conditions or strictly anaerobic conditions (80% N_2, 10% H_2, 10% CO_2) at $37\,^{\circ}C$ for 2 or 4 days, individual colonies were picked up and cultured for an additional 2 or 4 days at $37\,^{\circ}C$ in ABCM broth (Eiken Chemical) or EG agar plate. The isolated strains were collected into EG stock medium (10% glycerol) and stored at $-80\,^{\circ}C$. To identify the isolated strains, 16S rRNA gene sequences were determined. The 16S rRNA gene was amplified by colony-PCR using KOD FX (TOYOBO) and GeneAmp PCR System9700 (Applied Biosystems) using 16S rRNA gene-specific primer pairs: 8F (5′-AGAGTTTGATCMTGGCTCAG-3′) and 519R (5′-ATTACCGCGGCKGCTG-3′) or 1513R (5′-ACGGCTACCTTGTTACGACTT-3′). The amplification program consisted of one cycle at $98\,^{\circ}C$ for 2 min, followed by 40 cycles at $98\,^{\circ}C$ for 10 s, $57\,^{\circ}C$ for 30 s and $68\,^{\circ}C$ for 1 min 30 s. Each amplified DNA was purified from the reaction mixture using Illustra GFX PCR DNA and gel band purification kit (GE Healthcare). Sequence analysis was performed using BigDye Terminator V3.1 cycle sequencing kit (Applied Biosystems) and Applied Biosystems 3730xl DNA analyser (Applied Biosystems). The resulting sequences were compared with sequences in RDP database and genome database using BLAST to determine close species/strains.

Bacterial culture of isolated strains. The isolated strains of Clostridia and Erysipelotrichi were cultured in EG broth without horse blood under a strictly anaerobic condition (80% N_2, 10% H_2, 10% CO_2) at $37\,^{\circ}C$ in an anaerobic chamber (Coy Laboratory Products). To prepare the bacterial mixture, bacterial strains were individually grown in EG broth to confluence and mixed at equal amounts of media volume.

Scanning electron microscopy. Scanning electron microscopy was performed by Filgen, Inc., Japan. The proximal colon was removed from +23-mix mice, cut open longitudinally, prefixed with 2% glutaraldehyde in 0.1 M phosphate buffer (pH 7.4) for 24 h at $4\,^{\circ}C$, and then postfixed with 2% osmium tetroxide for 1 h at $4\,^{\circ}C$. Fixed samples were dehydrated for 5 min each in sequential baths of 50%, 70%, 90% and 100% ethanol, inserted into a critical point dryer until dry and coated with osmium in an OPC-80N osmium plasma coater (Filgen). Scanning electron micrographs were taken by a JEOL JSM-6320F instrument.

Measurement of organic acids. Organic acid concentrations in caecal contents were determined by gas chromatography-mass spectrometry (GC-MS). Caecal contents (10 mg) were disrupted using 3-mm zirconia/silica beads (BioSpec Products) and homogenized in extraction solution containing 100 µl of internal standard (100 µM crotonic acid), 50 µl of HCl and 200 µl of ether. After vigorous shaking using a Shakemaster neo (Bio Medical Science) at 1,500 r.p.m. for 10 min, homogenates were centrifuged at $1,000g$ for 10 min and then the top ether layer was collected and transferred into new glass vials. Aliquots (80 µl) of the ether extracts were mixed with 16 µl of N-tert-butyldimethylsilyl-N-methyltrifluoroacetamide (MTBSTFA). The vials were sealed tightly by screwing and heated at $80\,^{\circ}C$ for 20 min in a water bath, and left at room temperature for 48 h for derivatization. The derivatized samples were run through a 6890N Network GC System (Agilent Technologies) equipped with HP-5MS column (0.25 mm $\times$ 30 m $\times$ 0.25 µm) and 5973 Network Mass Selective Detector (Agilent Technologies). Pure helium (99.9999%) was used as a carrier gas and delivered at a flow rate of 1.2 ml min^{-1}.

The head pressure was set at 10 p.s.i. with split 10:1. The inlet and transfer line temperatures were 250 °C and 260 °C, respectively. The following temperature program was used: 60 °C (3 min), 60–120 °C (5 °C min^{-1}), 120–300 °C (20 °C min^{-1}). One microlitre quantity of each sample was injected with a run time of 30 min. Organic acid concentrations were quantified by comparing their peak areas with the standards.

Genome sequencing and gene prediction. The genome sequences of 17 T$_{reg}$-cell-inducing strains were determined by the whole-genome shotgun strategy using a 454GS FLX Ti or Ion PGM sequencer. Each 1–5 μg of the genomic DNA was sheared to obtain DNA fragments. Template DNA was prepared according to the supplier's protocol. The generated sequence data were assembled using Newbler v2.8 software to obtain the draft genome sequences. All genome sequence data were deposited in DDBJ BioProject ID: PRJDB521-543. Protein-encoding genes were predicted using MetaGeneAnnotator software[27]. Putative toxins and virulence factors were searched using the BLASTP program and virulence factor databases, VFDB (http://www.mgc.ac.cn/VFs/main.htm) and MvirDB (http://mvirdb.llnl.gov), with the e-value cutoff of 1.0×10^{-10}, the identity >30% and the length coverage >60%.

Phylogenetic tree. Sequences concatenated with genes encoding 26 ribosomal proteins (large subunit L10, L11, L14, L16, L17, L19, L20, L23, L24, L29, L31, L32, L35, L7/L12, and small subunit S10, S12, S13, S15, S16, S17, S20, S20, S21, S3, S4, S7, S8) predicted from the genomes of each strain were used to construct a phylogenetic tree. The sequences of other bacterial species used for the tree construction were obtained from the ribosomal multi-locus sequencing typing (MLST) database[28]. The calculation was performed using the MEGA v5.0 package and the neighbour-joining method with a bootstrap of 1,000 replicates.

Cognate antigen-driven T$_{reg}$ cell suppression assay. Preparation of antigens in caecal contents was performed as previously reported[15]. Caecal contents from germ-free mice or +17-mix mice were collected and suspended in PBS (500 mg ml^{-1}); they were then filtered through a 70-μm mesh, and autoclaved at 121 °C for 15 min. To prepare antigens of bacterial components, the 17 strains of Clostridia were cultured in vitro, mixed, washed and suspended with 1 ml PBS, and autoclaved at 121 °C for 20 min. CD11c$^+$ cells were isolated by FACSAriaIII from spleens of SPF C57BL/6 mice and pulsed for 1 h with 0.5 μM SIINFEKL OT-I peptide alone or in combination with either of 5 μM ISQAVHAAHAEINEAGR OT-II peptide, autoclaved caecal contents from +17-mix mice or germ-free mice (diluted 1:200), or autoclaved 17 strains of bacteria cultured in vitro (diluted 1:200). The antigen-pulsed CD11c$^+$ cells were plated at 5×10^4 per well in 96-well round-bottomed plates. CD8 T cells (T$_{eff}$ cells) were sorted from spleens of SPF OT-I mice by FACSAriaIII and added to the CD11c$^+$ cell-seeded plates at 5×10^4 per well. Then, CD4$^+$CD25$^+$ T cells (T$_{reg}$ cells) sorted from colonic lamina propria of +17-mix mice or from spleens of SPF OT-II mice were added to the culture at the indicated ratio of T$_{reg}$ to T$_{eff}$ cells. After 3 days, all cells were harvested, stained with anti-CD4 and anti-CD8 antibodies, and analysed by FACSAriaIII to enumerate the number of CD8 OT-I T cells.

Intestinal epithelial cell stimulation with caecal extracts and SCFAs. To prepare caecal extracts, frozen caecal contents from germ-free, +17-mix or SPF mice were thawed and well suspended in 4 volumes of sterile water. After centrifugation (5,000 r.p.m. for 15 min), transparent supernatants were collected, filtered through 0.22 μm filter and used as caecal extracts. In some experiments, caecal extracts were treated with proteinase K (2 mg ml^{-1}, 55 °C for 1 h; Roche) or nuclease that degrades all forms of DNA and RNA (125 unit ml^{-1}, 37 °C for 4 h; Thermo), and subsequently heated at 95 °C for 5 min to inactivate the enzymes. Human intestinal epithelial cell lines (HCT8, HT29, Caco2, T84 and Colo205) and a mouse

epithelial cell line (CMT93) were obtained from ATCC and maintained at 37 °C (5% CO2) in RPMI containing 10% heat-inactivated horse serum (Invitrogen). Cells were cultured at 1.5×10^5 cells in 150 μl medium in 48-well plates and stimulated with 4.5 μl caecal extract for 24 h. Human primary intestinal epithelial cells were obtained from Lonza and maintained at 33 °C (5% CO2) in SmGM-2 medium containing 10% FBS (Lonza) for 1–2 weeks (6×10^4 cells in 48-well plates). The medium was changed to 150 μl SmGM-2 containing 1% FBS before stimulation. Caecal extracts (4.5 μl) were added to the culture and incubated for 24 h. Culture supernatants were collected and the level of the active form of TGF-β1 (Promega), TNF-α (R&D) and IL-6 (R&D) was measured by ELISA. To stimulate epithelial cell lines with SCFAs, sodium salts of acetate, butyrate, propionate and isobutyrate were dissolved in PBS. SCFAs were added to the culture individually (final 0.5 mM) or in combination (final 0.5 mM each), and incubated for 24 h.

TNBS colitis. C57BL/6 SPF adult mice were orally inoculated with 17-mix or control PBS every 2 or 3 days for 3 weeks. 2,4,6-Trinitrobenzene sulphonic acid (TNBS)-induced colitis was induced by the intracolonic administration of 2.5 mg of TNBS (Sigma) in 50% ethanol into anaesthetized mice via a thin round-tip needle. The tip of the needle was inserted 4 cm proximal to the anal verge, and mice were held in a vertical position for 30 s after the injection. All the mice were observed daily and were killed on day 4 after TNBS administration. Colons were fixed with 4% paraformaldehyde, sectioned, and stained with haematoxylin and eosin. The degree of inflammation in the distal part of colon was graded from 0 to 4 as follows: 0, normal; 1, ulcer with cell infiltration limited to the mucosa; 2, ulcer with limited cell infiltration in the submucosa; 3, focal ulcer involving all layers of the colon; 4, multiple lesions involving all layers of the colon, or necrotizing ulcer larger than 1 mm in length.

Allergic diarrhoea. BALB/c SPF adult mice were primed by subcutaneous injection with 1 mg of OVA (Fraction V; Sigma) in 100 μl of Complete Freund Adjuvant (CFA, DIFCO). One week after priming, mice were given 50 mg of OVA dissolved in 200 μl of PBS by intra-gastric administration three times per week. 17-mix or control PBS was orally administered to mice every 2 or 3 days for the entire period of the experiments. Diarrhoea was monitored visually 1 h after each oral OVA challenge. Diarrhoea was scored as follows: 0, normal faeces (solid); 1, moist faeces (semi-solid); 2, mild diarrhoea (loose); and 3, severe diarrhoea (watery). Serum was collected from the cheek vein 1 h after the last OVA challenge and OVA-specific IgE levels were measured by ELISA (Chondrex).

Adoptive CD4$^+$CD45RB^{hi} T-cell transfer model of colitis. Germ-free CB.17 SCID mice were orally inoculated with SPF faeces together with or without 17-mix of Clostridia. One week later, exGF SCID mice received 4×10^5 CD4$^+$CD45RB^{hi} T cells by intraperitoneal injection. Naive CD4$^+$CD45RB^{hi} T cells were isolated from spleens of SPF BALB/c mice by FACS sorting. All the mice were observed daily and were killed on day 14 after T-cell transfer.

25. Morita, H. et al. An improved isolation method for metagenomic analysis of the microbial flora of the human intestine. Microbes Environ. **22,** 214–222 (2007).
26. Kim, S. W. et al. Robustness of gut microbiota of healthy adults in response to probiotic intervention revealed by high-throughput pyrosequencing. DNA Res. **20,** 241–253 (2013).
27. Noguchi, H., Taniguchi, T. & Itoh, T. MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. DNA Res. **15,** 387–396 (2008).
28. Jolley, K. A. et al. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. Microbiology **158,** 1005–1015 (2012).

LETTER

Pyrimidine homeostasis is accomplished by directed overflow metabolism

Marshall Louis Reaves[1,2], Brian D. Young[1,2], Aaron M. Hosios[1,2], Yi-Fan Xu[1,3] & Joshua D. Rabinowitz[1,2,3]

Cellular metabolism converts available nutrients into usable energy and biomass precursors. The process is regulated to facilitate efficient nutrient use and metabolic homeostasis. Feedback inhibition of the first committed step of a pathway by its final product is a classical means of controlling biosynthesis[1-4]. In a canonical example, the first committed enzyme in the pyrimidine pathway in *Escherichia coli* is allosterically inhibited by cytidine triphosphate[1,4,5]. The physiological consequences of disrupting this regulation, however, have not been previously explored. Here we identify an alternative regulatory strategy that enables precise control of pyrimidine pathway end-product levels, even in the presence of dysregulated biosynthetic flux. The mechanism involves cooperative feedback regulation of the near-terminal pathway enzyme uridine monophosphate kinase[6]. Such feedback leads to build-up of the pathway intermediate uridine monophosphate, which is in turn degraded by a conserved phosphatase, here termed UmpH, with previously unknown physiological function[7,8]. Such directed overflow metabolism allows homeostasis of uridine triphosphate and cytidine triphosphate levels at the expense of uracil excretion and slower growth during energy limitation. Disruption of the directed overflow regulatory mechanism impairs growth in pyrimidine-rich environments. Thus, pyrimidine homeostasis involves dual regulatory strategies, with classical feedback inhibition enhancing metabolic efficiency and directed overflow metabolism ensuring end-product homeostasis.

The metabolic network of *E. coli* consists of approximately 1,000 metabolites connected by around 2,000 enzyme-catalysed reactions[9]. Control of metabolite concentrations and fluxes occurs through the regulation of enzyme concentrations, activities and substrate occupancies. Metabolic control analysis provides a systematic framework for investigating the impact of particular enzymes on cellular metabolic activities[10-13]. Studies modulating the concentrations of enzymes suggest that control of metabolic flux is frequently distributed across multiple enzymes, with demand for end product often having a key role in controlling biosynthetic fluxes[2,14-16].

Consistent with distributed flux control, *de novo* pyrimidine biosynthesis has been reported to be regulated both at the first committed pathway step, catalysed by aspartate transcarbamoylase (ATCase), and the previous step, catalysed by carbamoyl phosphate synthetase (CPSase), which also feeds arginine biosynthesis. The *E. coli* ATCase enzyme complex, which consists of six catalytic and six regulatory subunits, is subject to feedback inhibition by the pyrimidine end products uridine triphosphate (UTP) and more strongly cytidine triphosphate (CTP), and is activated by ATP[1,5,17-19]. Its allosteric regulation provided one of the first examples of feedback inhibition[3,4]. CPSase is feedback inhibited by the pyrimidine intermediate uridine monophosphate (UMP)[20].

To explore the physiological relevance of ATCase and CPSase allostery, we created strains dysregulated for feedback control of ATCase (*ΔpyrI*), CPSase (*carB** (*carB(S948F)*)[21]) or both (*ΔpyrI carB**) (Fig. 1a). We then analysed the metabolite concentrations in these strains by liquid chromatography–mass spectrometry (LC–MS)-based metabolomics. We

anticipated that such strains would have increased levels of pyrimidine nucleotide triphosphates (NTPs)[2,11,16], the pathway's terminal products; however, UTP and CTP levels were steady in the absence of feedback control. Instead, the only notable change we observed was markedly increased uracil levels (Fig. 1b–d). To understand the robustness of this pathway, we also conducted transcriptome analysis. Rather than compensatory downregulation of pyrimidine biosynthetic genes in the *ΔpyrI carB** strain, we observed modest upregulation. In addition, we observed enhanced expression of genes involved in arginine synthesis (which also requires CPSase) and of the Rut pathway (a recently discovered uracil-degradation pathway that is induced by uracil) (Supplementary Fig. 1)[22].

To assess the homeostatic capacity of pyrimidine metabolism in response to pyrimidine intermediate addition, we switched *E. coli* grown on minimal media to media containing orotate or uracil. Such pyrimidine upshift was sufficient to activate feedback inhibition of pyrimidine synthesis in the wild type, as evidenced by reduced *N*-carbamoyl-aspartate and dihydroorotate concentrations within 5 min; this decrease did not occur in the *ΔpyrI carB** strain (Fig. 1e and Supplementary Fig. 2). In addition, pyrimidine upshift led to markedly increased uracil, and in the case of orotate addition, also UMP, and these increases were observed in all genetic backgrounds (Fig. 1e). Moreover, in both the presence and absence of feedback control, there were only minor increases in UTP and CTP.

Given the end-product homeostasis in the doubly feedback-dysregulated strain, even upon pyrimidine upshift, we sought to confirm that the *ΔpyrI carB** strain is indeed defective in *de novo* pyrimidine biosynthetic regulation. To this end, we measured the incorporation of isotopically labelled uracil or orotate into UTP and CTP. Similar to the wild-type strain, the *ΔpyrI carB** strain imported the labelled intermediates and incorporated them into end products. The residual production of unlabelled end products, indicative of persistent *de novo* synthesis, however, was higher in the *ΔpyrI carB** strain (Fig. 1f and Supplementary Fig. 2). This confirms that feedback regulation is functionally impaired.

If the feedback-regulation mechanisms facilitated superior pyrimidine homeostasis, perhaps too subtle to be detected by our LC–MS methods, we expected that the feedback-dysregulated strain would have a growth defect relative to its wild-type parent. The *ΔpyrI carB** strain did not, however, exhibit impaired growth in various nutrient conditions, including rich medium, minimal media, media enriched in nucleotides or amino acids, media limited for nitrogen or phosphate, or periodic switches between these conditions (Fig. 1g). We did, however, detect a modest (~10%) growth advantage for the wild-type strain under anoxic conditions and in the presence of the uncoupler 2,4-dinitrophenol.

Slower growth only during energy limitation suggests that the feedback-dysregulated strain engages in a chronic, energy-wasting process. Given the observed uracil excretion, a likely candidate for this inefficient process is the degradation of UMP to uracil. Indeed, orotate addition results in both higher UMP and higher uracil (Fig. 1e), and when the orotate is

[1]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544, USA. [2]Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA. [3]Department of Chemistry, Princeton University, Princeton, New Jersey 08544, USA.
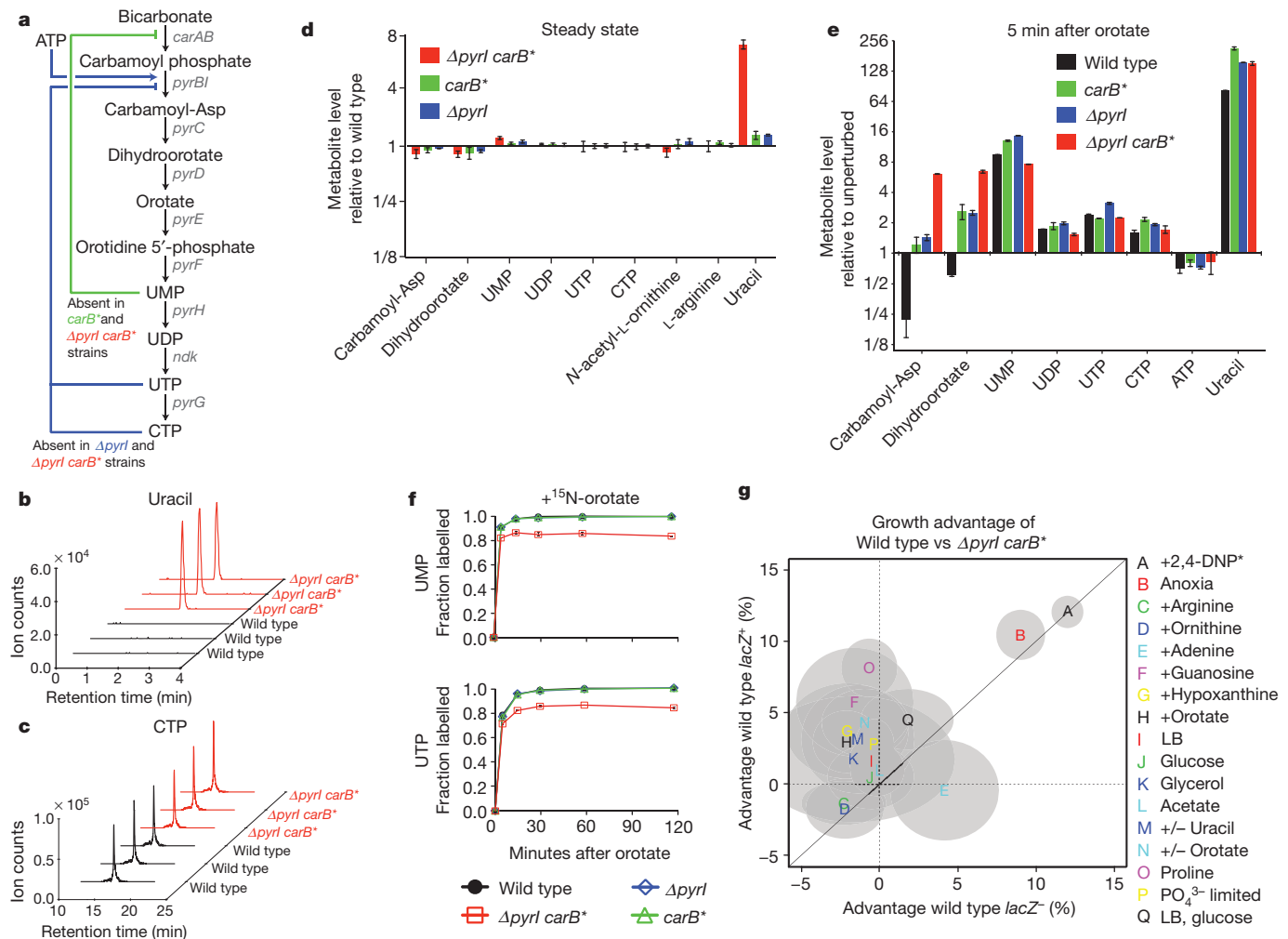
Figure 1 | Increased pyrimidine flux triggers overflow to uracil.
a, Canonical pyrimidine regulatory schematic. Carbamoyl phosphate is an intermediate in both pyrimidine and arginine synthesis. Carbamoyl aspartate is committed to pyrimidine synthesis. Carbamoyl phosphate synthetase (*carAB*) is feedback inhibited by UMP. Aspartate transcarbamoylase (*pyrBI*) is feedback inhibited by UTP and CTP and activated by ATP. **b**, **c**, Extracted ion chromatograms showing uracil (**b**) and CTP (**c**) levels in wild-type (black) and feedback-defective (*ΔpyrI carB**; red) strains. **d**, Metabolite fold changes relative to wild type in *ΔpyrI*, *carB** and *ΔpyrI carB** strains. Error bars denote ± standard error (n = 6). **e**, Metabolite fold changes at 5 min after addition of orotate. Fold changes were computed relative to un-supplemented controls. Error bars denote ± standard deviation (n = 2–3). Time course data appear in Supplementary Fig. 2. **f**, Fraction of UMP and UTP derived from endogenous

sources and from exogenously added ^{15}N-orotate. Error bars denote ± standard deviation (n = 2). The *ΔpyrI* (blue) and wild-type (black) lines lie under the *carB** (green) line. **g**, Competitive growth advantage of wild-type versus *ΔpyrI carB** appears selectively under energy limitation. Competitions were performed in indicated media with *lacZ*⁻ marker in wild-type and in feedback-dysregulated strain (wild-type *lacZ*⁺) to control for effect of marker on growth. Calculations and experimental details are described in Methods. In brief, media were glucose–ammonia minimal media unless otherwise indicated. Alternative carbon and nitrogen sources were used in conditions K, L and O. + indicates supplements to the minimal media; +/− indicates alternating supplementation/removal of indicated nutrient every 8 h. Grey ellipses mark 95% confidence interval (n = 6–10).

[15]N-labelled, the extent of uracil isotope labelling mimics that of UMP (Supplementary Fig. 3). To elucidate the route from UMP to uracil, we knocked out the genes known to catalyse the interconversion of UMP, uridine and uracil[17]: *udk* (uridine + GTP → UMP + GDP), *upp* (uridine + PRPP → UMP + pyrophosphate (PPᵢ)) and *udp* (uridine + inorganic phosphate (Pᵢ) → ribose-1-P + uracil). Only the *udp* knockout reduced uracil accumulation, decreasing it to 17% of the level normally found in the feedback-dysregulated background (Fig. 2b). We were able to eliminate uracil production more fully (to 6%) by double deletion of *udp* and the cytosine deaminase *codA*. Following orotate addition, this double-deletion mutant feedback-dysregulated strain accumulates uridine instead of uracil (Fig. 2c and Supplementary Fig. 4).

There is no enzyme with a well-described physiological function of UMP degradation to uridine. This reflects the lack of a straightforward genetic screen for such enzymes. Although often annotated as reversible, the Udk and Upp reactions are thermodynamically unfavourable

for the degradation of UMP (both ΔG_{Udk} and $\Delta G_{Upp} = +22\,\mathrm{kJ\,mol^{-1}}$ based on $\Delta G^{\circ}_{Udk} = +19.5\,\mathrm{kJ\,mol^{-1}}$ and $\Delta G^{\circ}_{Upp} = +20\,\mathrm{kJ\,mol^{-1}}$ and [GTP] = 4.9 mM, [GDP] = 0.68 mM, [uracil] = 5.5 mM, [uridine] = 0.15 mM, [UMP] = 0.5 mM, [diphosphate] = ~1 mM and [PRPP] = 0.26 mM[23,24], and indeed, their genetic deletion does not prevent uracil accumulation (Fig. 2b). On the basis of published *in vitro* enzymology and homology[17], we identified ~15 putative UMP phosphatase, hydrolase and nucleotidase genes with unknown physiological functions and screened deletion mutants for uracil levels following orotate addition. We found two genes whose deletion led to lower uracil and higher UMP levels than the wild type (Fig. 2c), indicating a possible role in UMP degradation in conditions of pyrimidine excess. We propose that these genes, *nagD* and *surE*, be renamed *umpH* and *umpG*, respectively, to highlight their newly recognized activities in UMP degradation. UmpH and UmpG belong to the haloacid dehalogenase-like phosphatase family and SurE phosphatase families, respectively, whose members span taxa from bacteria to humans, but little is known
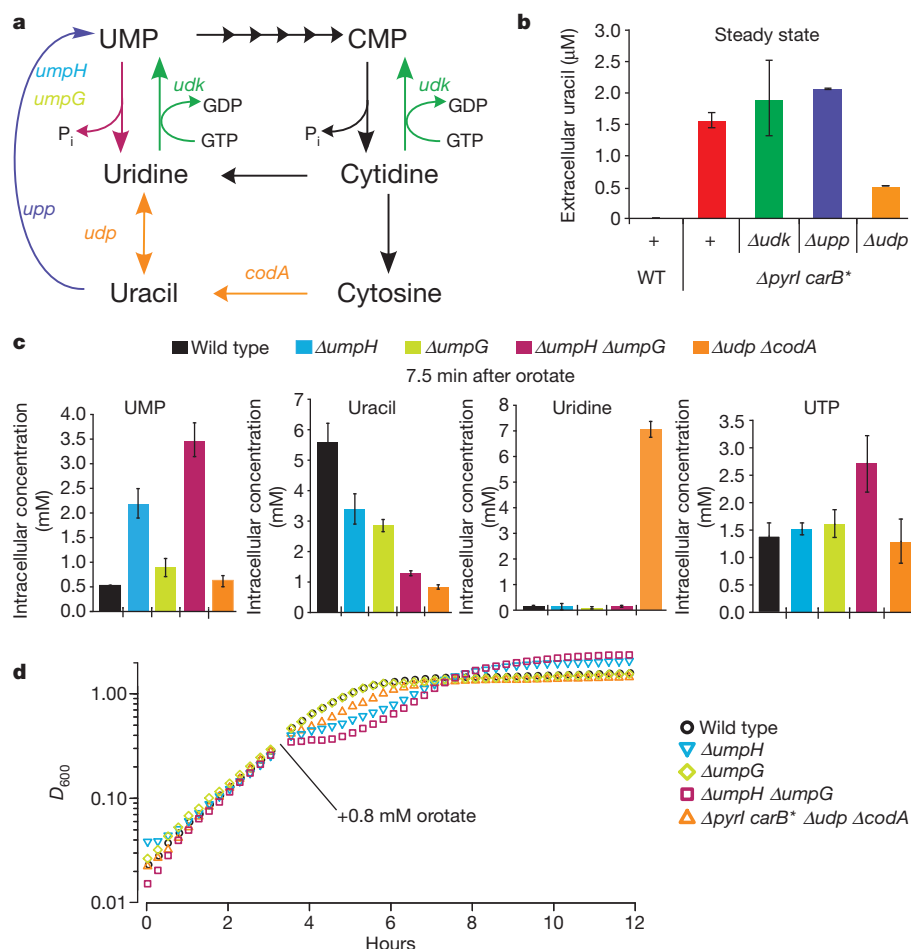
Figure 2 | Pyrimidine overflow pathway is initiated by catabolism of UMP by UmpH. a, Pathway schematic. **b**, Uracil excretion does not depend on the canonical pyrimidine interconversion enzymes Udk and Upp, but does require Udp. Excreted uracil accounts for approximately half of total uracil. Error bars mark standard deviation ($n = 2–3$). **c**, Uracil is produced by UMP degradation following orotate addition. UMP is degraded to uridine by UmpH (also known as NagD) and UmpG (also known as SurE), and uridine to uracil by Udp. Isotopic tracing from orotate to UMP, uridine and uracil appears in

Supplementary Fig. 3. Residual uracil in a $\Delta udp \Delta codA$ strain indicates at least one unannotated activity also makes minor contributions to uracil production. End-product levels are not perturbed following orotate addition by deletion of either $\Delta umpH$ or $\Delta umpG$ individually, but double deletion leads to increased UTP. Error bars mark $\pm$ standard deviation ($n = 3$). **d**, Inability to degrade UMP to uracil causes a growth defect upon orotate addition but increases final culture density (reproducible in two biologically independent experiments each of 6–8 technical replicates). D, attenuance.

about their physiological importance[7,8,25,26]. We observed that knockout of *umpH*, although not altering pathway end-product levels, impaired growth of *E. coli* upon orotate upshift, with the double deletion showing a stronger phenotype and end-product accumulation (Fig. 2c, d and Supplementary Fig. 4). Thus, UmpH, and to a lesser extent UmpG, function in a UMP-degradation pathway that is required for optimal growth in response to environmental pyrimidine intermediates.

The UmpH and UmpG phosphatases degrade UMP to uridine, which is further degraded by Udp to uracil and ribose-1′-phosphate (cytosine similarly liberated from CMP may be deaminated by CodA to uracil). This overflow pathway dissipates three high-energy phosphate bonds (ATP equivalents) and one NADPH per uracil excreted aerobically, increasing to five ATP and inducing production of ubiquinol under anoxia, which may be problematic (Supplementary Fig. 5). The most straightforward way to control flux through this overflow pathway is through the concentration of its substrate UMP. In all conditions where we observe uracil excretion, we also observe increased intracellular UMP. The normal steady-state UMP concentration (0.052 mM) is below the Michaelis constant (K_m) of UmpH for UMP (0.12 mM), but rises to eight times the K_m under conditions inducing directed overflow metabolism (Fig. 2c). This suggests that UMP accumulation has a central role in triggering the overflow flux.

We sought to identify the molecular underpinnings of UMP accumulation during growth in medium enriched in pyrimidine intermediates.

In analysing the regulatory architecture of the pathway, we noticed an additional potential feedback loop near the end of the pathway: UMP kinase (which catalyses the ATP-dependent phosphorylation of UMP to UDP), encoded by *pyrH* (Fig. 3a), is regulated *in vitro* by cooperative (ultrasensitive) inhibition by UTP (Hill-coefficient of 2.8, inhibition constant (K_i) of 154 μM)[6]. This inhibition can be partially overcome by GTP activation[6,27], and thus, like the regulation of ATCase by CTP and ATP, may function to achieve balance between pyrimidine and purine pools. Moreover, regulation at the end of the pathway would have the advantage of minimizing perturbations in UTP and CTP levels in response to alterations in any upstream pathway substrate (for example, not only of the *de novo* pathway, but also uracil, orotate, or their nucleosides).

To test the importance of the cooperative inhibition of UMP kinase by UTP, we created strains expressing mutant forms of UMP kinase with reduced sensitivity to UTP: the N72A mutant has a higher K_i (373 μM for UTP rather than 154 μM), and the D93A has both a higher K_i (332 μM) and less cooperative inhibition (Hill coefficient 1.6 rather than 2.8). Because UMP kinase catalyses an essential reaction, we first transformed wild-type cells with low-copy plasmids (pACYC) carrying a natively expressed *pyrH* allele (*pyrH*[WT], N72A or D93A) before knocking out genomic *pyrH* by transduction. We then assayed the growth and metabolic response of these strains to uracil and orotate. Although all three strains grew normally in the
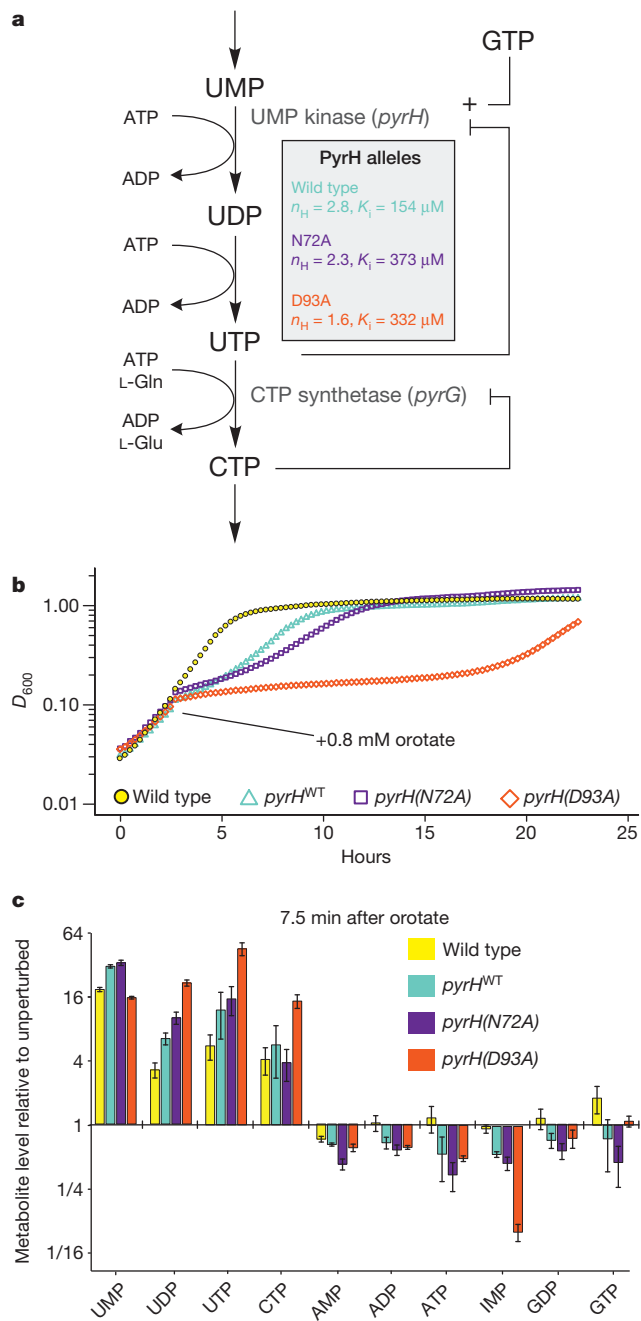
a

Figure 3 | Cooperative inhibition of UMP kinase by UTP maintains end-product homeostasis. Variant *pyrH* alleles were expressed from its native promoter on low-copy plasmid (pACYC) with genomic *pyrH* removed. **a**, Schematic of downstream regulatory events in pyrimidine metabolism. Wild-type UMP kinase (PyrH) is feedback inhibited by UTP in a switch-like manner (high degree of cooperativity). Allosteric parameters of *pyrH* alleles appear in shaded box, with D93A lacking the switch-like behaviour. n_H, Hill coefficient. **b**, Altered expression or regulation of UMP kinase following orotate addition impairs growth. Defects also occur in *pyrH*^WT^/*pyrH* diploid strains (Supplementary Fig. 6). **c**, Metabolite fold changes upon orotate addition in strains with altered UMP kinase expression or allosteric regulation reveal defects in both pyrimidine and purine homeostasis. Error bars mark standard deviation ($n = 3$).

absence of pyrimidines and upon uracil addition (Supplementary Fig. 6), orotate addition inhibited growth of all three strains, particularly those with impaired inhibition of UMP kinase by UTP (Fig. 3b). The growth defect of the *pyrH*^WT^ (pACYC::*pyrH*^WT^ *ΔpyrH*) strain was presumably due to mild overexpression of wild-type UMP kinase, as inducible overexpression led to the same response (pAC24N::*pyrH*^WT^ + 0.05 μM

isopropyl β-D-1-thiogalactopyranoside (ITPG); Supplementary Fig. 6). Thus, proper control of UMP kinase protein levels is required for an optimal response to pyrimidine upshift. The cooperative feedback inhibition of UMP kinase by UTP and CTP is also important, as impairment of such regulation led to a profound growth defect upon orotate addition (pACYC::*pyrH(D93A) ΔpyrH*, Fig. 3b).

We also detected changes to the nucleotide pools of these *pyrH* mutants in response to orotate addition. In the wild type, orotate addition leads to UMP accumulation without major changes in NTP levels (Fig. 3c). Lack of proper UMP kinase regulation, however, results in increased UTP and decreased ATP. Thus, dysregulated pyrimidine metabolism saps purines, either by increasing their usage (for example, because higher pyrimidine NTPs leads to increased ribosomal RNA biosynthetic rates), decreasing their synthesis (for example, by depleting required substrates), or the combination of these factors.

We designate the general mechanism, whereby feedback inhibition of a downstream pathway step leads to excretion of a pathway intermediate or by-product, directed overflow metabolism (Fig. 4a). Such overflow is triggered by excessive biosynthetic pathway flux and is carried out by a degradation pathway sensitive to levels of an accumulating biosynthetic intermediate. It is analogous to overflow in central carbon metabolism, wherein excessive sugar catabolism (typically via glycolysis) leads to build-up of pyruvate[28], which may be excreted as lactate, ethanol or acetate, depending on the organism. In the case of lactate excretion in humans, inhibition of pyruvate dehydrogenase (by high NADH or inhibitory phosphorylation[29]) has a similar role to the inhibition of UMP kinase in the pyrimidine pathway: it forms a choke point that instead directs the enzyme's normal substrate towards by-product formation and excretion (Fig. 4b). In the case of pyrimidine biosynthesis, the cooperative inhibition of UMP kinase by UTP renders the directed overflow mechanism exquisitely precise in controlling UTP and CTP levels.

Thus, pyrimidine homeostasis involves two strategies for regulation. The canonical feedback architecture contributes to metabolic efficiency by decreasing unnecessary *de novo* flux. Directed overflow metabolism provides end-product homeostasis by diverting excess flux to uracil, thereby ensuring end-product homeostasis in response to altered availability of the full range of pathway substrates and intermediates. These
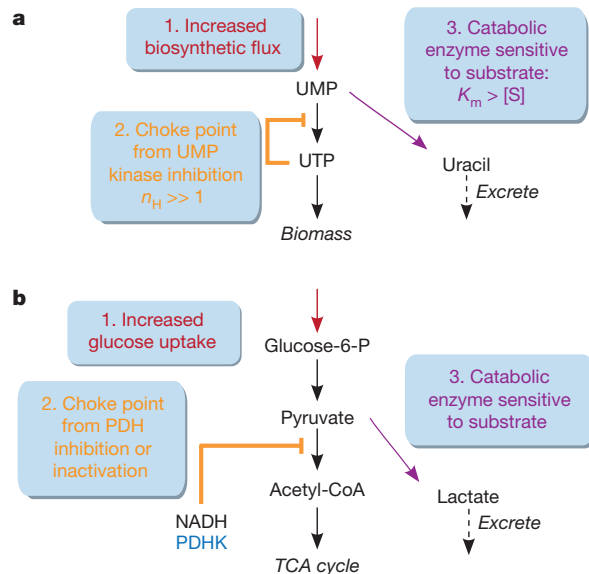

Figure 4 | Directed overflow metabolism in biosynthesis is analogous to central carbon overflow metabolism. a, Schematic of directed overflow metabolism as a biosynthetic regulatory mechanism. **b**, Schematic of overflow in central carbon metabolism, using the Warburg effect as a canonical example. PDH, pyruvate dehydrogenase; PDHK, pyruvate dehydrogenase kinase (which catalyses inhibitory phosphorylation of PDH); TCA, tricarboxylic acid.

upstream and downstream regulatory mechanisms work in concert to balance speed, efficiency and robustness.

METHODS SUMMARY

E. coli (parent strain NCM3722 (ref. 30)) were prepared for metabolite measurement using a filter culture technique: cell-laden nitrocellulose filters were grown on agarose plates. Pyrimidine upshift was accomplished by transferring filters to plates enriched with indicated supplements. Metabolism was quenched and metabolites concomitantly extracted by placing filters into −20 °C solvent (40% methanol, 40% acetonitrile and 20% water with 0.1 M formic acid). This solvent mixture reliably extracts nucleotides, nucleosides and bases without their interconversion or degradation. Negative mode LC–MS and LC–MS/MS measurements were performed as previously described[23]. Deletion strains were created by P1 transduction from the Keio deletion collection. Stable strains carrying *carB** were generated by electroporation and lambda Red-mediated recombination of PCR-amplified S948F into a *ΔcarB::kan* strain (where *kan* denotes kanamycin resistance) followed plating and selection for prototrophs on minimal media. Growth was assayed by absorbance at 600 nm in a 96-well format using a Biotek Synergy II plate reader. Competitive growth advantage was assessed by co-culture of a marked *lacZ⁻* (*ΔlacZ::kan*) and an unmarked *lacZ⁺* strain, determination of relative cell numbers after competitive growth by plating on MacConkey agar containing 1% lactose, and regression analysis of wild type to *ΔpyrI carB** ratios. Unless otherwise indicated, growth media for metabolic studies consisted of Gutnick minimal media with 0.4% glycerol and 10 mM NH₄Cl with triple-washed ultra-pure agarose and 0.8 mM orotate as needed.

Full Methods and any associated references are available in the online version of the paper.

1. Gerhart, J. C. & Pardee, A. B. The enzymology of control by feedback inhibition. *J. Biol. Chem.* **237,** 891–896 (1962).
2. Savageau, M. A. Optimal design of feedback-control by inhibition — dynamic considerations. *J. Mol. Evol.* **5,** 199–222 (1975).
3. Umbarger, H. E. Evidence for a negative-feedback mechanism in the biosynthesis of isoleucine. *Science* **123,** 848–849 (1956).
4. Pardee, A. B. & Yates, R. A. Control of pyrimidine biosynthesis in *Escherichia coli* by a feed-back mechanism. *J. Biol. Chem.* **221,** 757–770 (1956).
5. Kantrowitz, E. R. Allostery and cooperativity in *Escherichia coli* aspartate transcarbamoylase. *Arch. Biochem. Biophys.* **519,** 81–90 (2012).
6. Meyer, P. *et al.* Structural and functional characterization of *Escherichia coli* UMP kinase in complex with its allosteric regulator GTP. *J. Biol. Chem.* **283,** 36011–36018 (2008).
7. Kuznetsova, E. *et al.* Genome-wide analysis of substrate specificities of the *Escherichia coli* haloacid dehalogenase-like phosphatase family. *J. Biol. Chem.* **281,** 36149–36161 (2006).
8. Proudfoot, M. *et al.* General enzymatic screens identify three new nucleotidases in *Escherichia coli.* Biochemical characterization of SurE, YfbR, and YjjG. *J. Biol. Chem.* **279,** 54687–54694 (2004).
9. Orth, J. D. *et al.* A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Mol. Syst. Biol.* **7,** 535 (2011).
10. Fell, D. *Understanding the Control of Metabolism* (Portland Press, 1997).
11. Kacser, H., Burns, J. A. & Fell, D. A. The control of flux. *Biochem. Soc. Trans.* **23,** 341–366 (1995).
12. Heinrich, R. & Rapoport, T. A. Linear steady-state treatment of enzymatic chains. General properties, control and effector strength. *Eur. J. Biochem.* **42,** 89–95 (1974).
13. Crabtree, B. & Newsholme, E. A. The derivation and interpretation of control coefficients. *Biochem. J.* **247,** 113–120 (1987).
14. Small, J. R. & Kacser, H. Responses of metabolic systems to large changes in enzyme-activities and effectors. 1. The linear treatment of unbranched chains. *Eur. J. Biochem.* **213,** 613–624 (1993).
15. Kell, D. B. & Westerhoff, H. V. Metabolic control theory: its role in microbiology and biotechnology. *FEMS Microbiol. Rev.* **39,** 305–320 (1986).
16. Hofmeyr, J. H. S. & Cornish-Bowden, A. Quantitative assessment of regulation in metabolic systems. *Eur. J. Biochem.* **200,** 223–236 (1991).
17. Keseler, I. M. *et al.* EcoCyc: a comprehensive database of *Escherichia coli* biology. *Nucleic Acids Res.* **39,** D583–D590 (2011).
18. Peterson, A. W., Cockrell, G. M. & Kantrowitz, E. R. A second allosteric site in *Escherichia coli* aspartate transcarbamoylase. *Biochemistry* **51,** 4776–4778 (2012).
19. Wild, J. R., Loughrey-Chen, S. J. & Corder, T. S. In the presence of CTP, UTP becomes an allosteric inhibitor of aspartate transcarbamoylase. *Proc. Natl Acad. Sci. USA* **86,** 46–50 (1989).
20. Anderson, P. M. & Meister, A. Control of *Escherichia coli* carbamyl phosphate synthetase by purine and pyrimidine nucleotides. *Biochemistry* **5,** 3164–3169 (1966).
21. Delannay, S. *et al.* Serine 948 and threonine 1042 are crucial residues for allosteric regulation of *Escherichia coli* carbamoylphosphate synthetase and illustrate coupling effects of activation and inhibition pathways. *J. Mol. Biol.* **286,** 1217–1228 (1999).
22. Loh, K. D. *et al.* A previously undescribed pathway for pyrimidine catabolism. *Proc. Natl Acad. Sci. USA* **103,** 5114–5119 (2006).
23. Bennett, B. D. *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli. Nature Chem. Biol.* **5,** 593–599 (2009).
24. Flamholz, A., Noor, E., Bar-Even, A. & Milo, R. eQuilibrator–the biochemical thermodynamics calculator. *Nucleic Acids Res.* **40,** D770–D775 (2012).
25. Bianchi, V. & Spychala, J. Mammalian 5′-nucleotidases. *J. Biol. Chem.* **278,** 46195–46198 (2003).
26. Tremblay, L. W., Dunaway-Mariano, D. & Allen, K. N. Structure and activity analyses of *Escherichia coli* K-12 NagD provide insight into the evolution of biochemical function in the haloalkanoic acid dehalogenase superfamily. *Biochemistry* **45,** 1183–1193 (2006).
27. Bucurenci, N. *et al.* Mutational analysis of UMP kinase from *Escherichia coli. J. Bacteriol.* **180,** 473–477 (1998).
28. Sauer, U. & Eikmanns, B. J. The PEP–pyruvate–oxaloacetate node as the switch point for carbon flux distribution in bacteria. *FEMS Microbiol. Rev.* **29,** 765–794 (2005).
29. Roche, T. E. & Hiromasa, Y. Pyruvate dehydrogenase kinase regulatory mechanisms and inhibition in treating diabetes, heart ischemia, and cancer. *Cell. Mol. Life Sci.* **64,** 830–849 (2007).
30. Soupene, E. *et al.* Physiological studies of *Escherichia coli* strain MG1655: growth defects and apparent cross-regulation of gene expression. *J. Bacteriol.* **185,** 5611–5626 (2003).

METHODS

Strains. The wild-type *E. coli* K-12 strain NCM3722 was used in this study as wild type and as parent for all strains in this study because it lacks the *rph*-1 mutation carried by MG1655 that causes pyrimidine pseudo-auxotrophy from reduced expression of orotate phosphoribosyltransferase (*pyrE*)[30]. Single-gene deletion mutants of NCM3722 were generated by P1 transduction[31] of deletion alleles with kanamycin resistance (*kan*) cassettes from the Keio collection[32] and verified using PCR with gene-specific primers. Multiple gene deletions were similarly created serially using the FLP helper plasmid system (pCP20), which was also used to eliminate kanamycin resistance cassettes to produce scarred deletions[33].

The *carB(S948F)* mutant was generated using the lambda-Red recombinase system[33]. A plasmid-borne *S948F* mutant allele of *carB*[21] was amplified from plasmid by PCR (*carB* forward primer, 5′-CGCATAAATCCCTGTTCGAC-3′; *carB* reverse primer, 5′-CCATTCGGCGATTAACAAGT-3′), gel-purified and used to transform a NCM3722 Δ*carB::kan* strain carrying pKD46 and induced with arabinose as described previously[32]. Recombinants were selected on minimal media for pyrimidine and arginine prototrophy and subsequently cured of pKD46. Following screening for loss of all antibiotic resistance (ampicillin and kanamycin), the *carB(S948F)* (*carB**) mutation was confirmed by sequencing.

Strains natively expressing variant *pyrH* alleles in a Δ*pyrH::kan* background were also created using the lambda-Red recombinase system. The pACYC::*pyrH* is the ligation product of XbaI-digested, CAP-treated pACYC184 (all New England Biolabs) with XbaI-digested wild-type *pyrH* generated by PCR from NCM3722 genomic template using primer set (*pyrH* forward primer, 5′-ACTCTAGACCAA TGCAAAACCCGTCTAT-3′; *pyrH* reverse primer 5′-GATCTAGAACTTACG CGGAATCTTACCC-3′) (XbaI sites are underlined). Variant alleles were produced through site-directed mutagenesis (Genewiz) and verified by sequencing. NCM3722 carrying plasmid pACYC::*pyrH* and pKD46 was transformed with a *kan* cassette flanked by *pyrH* homology regions (created by PCR of a Keio collection strain with PAGE-purified primer set (forward, 5′-TTGTAAATTCAGCTA ACCCTTGTGGGGCTGCGCTGAATTCCGGGGATCCGTCGACC-3′; reverse, 5′-ACCAAACTGCCTGCAACAATAACGCCTTATAACCAGTGTAGGCTGG AGCTGCTTCG-3′)) to create a genomic Δ*pyrH::kan* allele. To prevent suppressing mutations, the Δ*pyrH::kan* allele was introduced by P1 transduction to strains carrying *pyrH* alleles on the pACYC plasmid and thereafter strains retained plasmid born resistances in absence of antibiotic.

Strains expressing *pyrH* from IPTG-inducible promoter on the pAC24N were created by transformation with *pAC24N::pyrH* from the Aska collection after removal of green fluorescent protein by NotI digestion and self-ligation[34]. Variant alleles were produced through site-directed mutagenesis (Genewiz) and verified by sequencing. Low-level expression was induced with IPTG as indicated.

Media and bacterial culture. Gutnick glucose minimal medium ('Glucose', Gmin1) refers to a salts mixture[35] with a 0.4% (w/v) glucose carbon source and a 10 mM ammonium chloride nitrogen source unless otherwise stated. 'Glycerol' and 'acetate' refer to the identical Gutnick with 0.4% glycerol or 0.4% sodium acetate, respectively, substituted for glucose. Strains were also grown in media supplemented (+) with one of the following: adenine, arginine, guanosine, hypoxanthine, ornithine, uracil or 2,4-dinitrophenol (all 1 mM). Media with alternative nitrogen sources contained Gutnick salts, 0.4% glycerol and 10 mM total concentration of usable nitrogen. Phosphate-limited MOPS (TekNova) medium was prepared with 0.4% glycerol, 10 mM ammonium chloride and 0.5 mM monobasic potassium phosphate[36]. Anoxia experiments were conducted in Gmin1 media at 37 °C inside a chamber containing approximately 90% nitrogen gas, 5% carbon dioxide and 5% hydrogen. Oxygen levels were maintained at 0 p.p.m. and rarely exceeded 100 p.p.m. +/− indicates alternating every 8 h between glycerol minimal media and enriched minimal media as indicated above. 'LB, glucose' indicates cells were alternated between LB and Gmin1 media every 8 h. Difco MacConkey agar (Becton, Dickinson) was prepared according to the manufacturer's instructions with 1% lactose as carbon source.

Metabolite measurements. Cells were cultured and extracted at −20 °C 40:40:20 methanol:acetonitrile:water with 0.1 M formic acid[37] using the filter culture method described previously[38]. Low-molecular-weight metabolites were quantitated using both LC–MS[39] and LC–MS/MS[40] at both steady state[23] and following addition of exogenous pyrimidines, including unlabelled and isotopically labelled uracil and orotate. Absolute quantification of uracil and UMP was computed as in ref. 23. LC–MS(/MS) data were analysed using MAVEN software[41]. U-[^{15}N]-tracers (for example, uracil and orotate) were purchased from Cambridge Isotopes Laboratories.

Competition assays. The *lacZ*− (Δ*lacZ::kan*) derivatives of wild-type and Δ*pyrI carB** strains were prepared as described above. After overnight culture in a 50:50 mix of medium for competition and Gmin1, *lacZ*+ and *lacZ*− strains were combined to equal proportions (based on attenuance D_{600}) and were diluted to an OD_{600} of 0.02 in medium for competition to create two paired cultures: wild type *lacZ*− with Δ*pyrI carB** *lacZ*+ and wild type *lacZ*+ with Δ*pyrI carB** *lacZ*−. Each culture pair was grown at 37 °C and generally diluted every 8 h (12 h for media with doubling time over 120 min) into fresh medium to D_{600} of 0.02. Some cultures were diluted into alternating media as indicated above.

The ratio of wild type to Δ*pyrI carB** of the culture was determined by regular plating on MacConkey agar plates with 1% lactose (6–10 replicates). Cultures were diluted to ensure that approximately 100 colonies formed on each plate.

Growth advantage was computed by a linear regression of the ratio of wild-type and mutant strains. The ratio of wild-type to mutant cells in time, *t*, can be described as $R(t) = R_0 \times 2^{t*(1/W-1/M)}$, where R_0 is the initial ratio, and *W* and *M* are the doubling times for the wild type and the mutant, respectively. Performing a linear regression of this equation (R software, http://cran.r-project.org) in logarithmic space allows for computation of the growth advantage.

For 2,4-dinitrophenol, advantage was similarly computed from the growth rate of each strain (*lacZ*+) individually in culture and plotted along the diagonal.

Microarrays. RNA purification, labelling, microarray measurement and data processing were performed as described in ref. 42. In brief, overnights of the wild-type and Δ*pyrI carB** mutant were prepared as above in Gmin2 and grown to mid-exponential phase (D_{600} of 0.4) and then mixed with RNAprotect Bacteria Reagent according to manufacturer's specifications (Qiagen). After 5 min at 25 °C, the mixture was pelleted for 10 min at 5,000*g*, and RNA was extracted using the Total RNA Purification kit (Norgen Biotek). Purified total RNA was treated for 30 min using *E. coli* poly(A) polymerase (New England Biolabs). RNA from the wild-type and Δ*pyrI carB** strains were labelled with Cy5 and Cy3, respectively, using the Two-Colour Low Input Quick Amp Labelling Kit (Agilent). Samples were hybridized to an *E. coli* Gene Expression Microarray (Agilent). Resulting data were analysed using iPAGE[43] and R software.

Growth assays. Absorbance measurements were obtained using a Biotek Synergy II plate reader (Biotek, Winooski, VT) in 96-well format at 600 nm using clear, flat-bottom plates with two independent biological replicates each consisting of six to eight technical replicates. Cultures included 0.1% Tween-20 to prevent clumping and were sealed with air permeable membrane. To facilitate comparisons, some curves are aligned to D_{600} of 0.03.

31. Silhavy, T. J., Berman, M. L. & Enquist, L. W. *Experiments With Gene Fusions* (Cold Spring Harbor Press, 1984).
32. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2,** 2006.0008 (2006).
33. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97,** 6640–6645 (2000).
34. Kitagawa, M. *et al.* Complete set of ORF clones of *Escherichia coli* ASKA library (a complete Set of *E. coli* K-12 ORF archive): unique resources for biological research. *DNA Res.* **12,** 291–299 (2006).
35. Gutnick, D., Calvo, J. M., Klopotowski, T. & Ames, B. N. Compounds which serve as the sole source of carbon or nitrogen for *Salmonella typhimurium* LT-2. *J. Bacteriol.* **100,** 215–219 (1969).
36. Neidhardt, F. C., Bloch, P. L. & Smith, D. F. Culture medium for enterobacteria. *J. Bacteriol.* **119,** 736–747 (1974).
37. Rabinowitz, J. D. & Kimball, E. Acidic acetonitrile for cellular metabolome extraction from *Escherichia coli*. *Anal. Chem.* **79,** 6167–6173 (2007).
38. Bennett, B. D., Yuan, J., Kimball, E. H. & Rabinowitz, J. D. Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach. *Nature Protocols* **3,** 1299–1311 (2008).
39. Lu, W. *et al.* Metabolomic analysis via reversed-phase ion-pairing liquid chromatography coupled to a stand alone orbitrap mass spectrometer. *Anal. Chem.* **82,** 3212–3221 (2010).
40. Lu, W., Kimball, E. & Rabinowitz, J. D. A high-performance liquid chromatography-tandem mass spectrometry method for quantitation of nitrogen-containing intracellular metabolites. *J. Am. Soc. Mass Spectrom.* **17,** 37–50 (2006).
41. Clasquin, M. F., Melamud, E. & Rabinowitz, J. D. LC-MS data processing with MAVEN: a metabolomic analysis and visualization engine. *Curr. Protoc. Bioinformatics* **14,** 14.11 (2012).
42. Goodarzi, H. *et al.* Regulatory and metabolic rewiring during laboratory evolution of ethanol tolerance in *E. coli*. *Mol. Syst. Biol.* **6,** 378 (2010).
43. Goodarzi, H., Elemento, O. & Tavazoie, S. Revealing global regulatory perturbations across human cancers. *Mol. Cell* **36,** 900–911 (2009).

CORRECTIONS & AMENDMENTS

Corrigendum: Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions

Lars Guelen, Ludo Pagie, Emilie Brasset, Wouter Meuleman, Marius B. Faza, Wendy Talhout, Bert H. Eussen, Annelies de Klein, Lodewyk Wessels, Wouter de Laat & Bas van Steensel

In this Letter, Fig. 4g and Supplementary Fig. 5f depict an analysis of H3K9me3 (not H3K9me2) ChIP data. In addition, three sentences in the third paragraph of the 'Microarray data analysis' section of the Methods should read: "Gene expression data and H3K27me3 ChIP data are from Tig3 lung fibroblasts[18]. H3K9me3 ChIP data are from human fetal lung fibroblasts[30]… H3K4me2 and PolII ChIP data are from WI38 human lung fibroblasts[29]."

Corrigendum: Immune surveillance by CD8$\alpha\alpha^+$ skin–resident T cells in human herpes virus infection

Jia Zhu, Tao Peng, Christine Johnston, Khamsone Phasouk, Angela S. Kask, Alexis Klock, Lei Jin, Kurt Diem, David M. Koelle, Anna Wald, Harlan Robins & Lawrence Corey

In Fig. 2a of our Letter, the label 'CXCR7' should be 'CXCR6', and the label 'CXCR8' should be 'CXCR7'. This error has been corrected in the HTML and PDF versions of the paper.

Erratum: Integrated genomic characterization of endometrial carcinoma

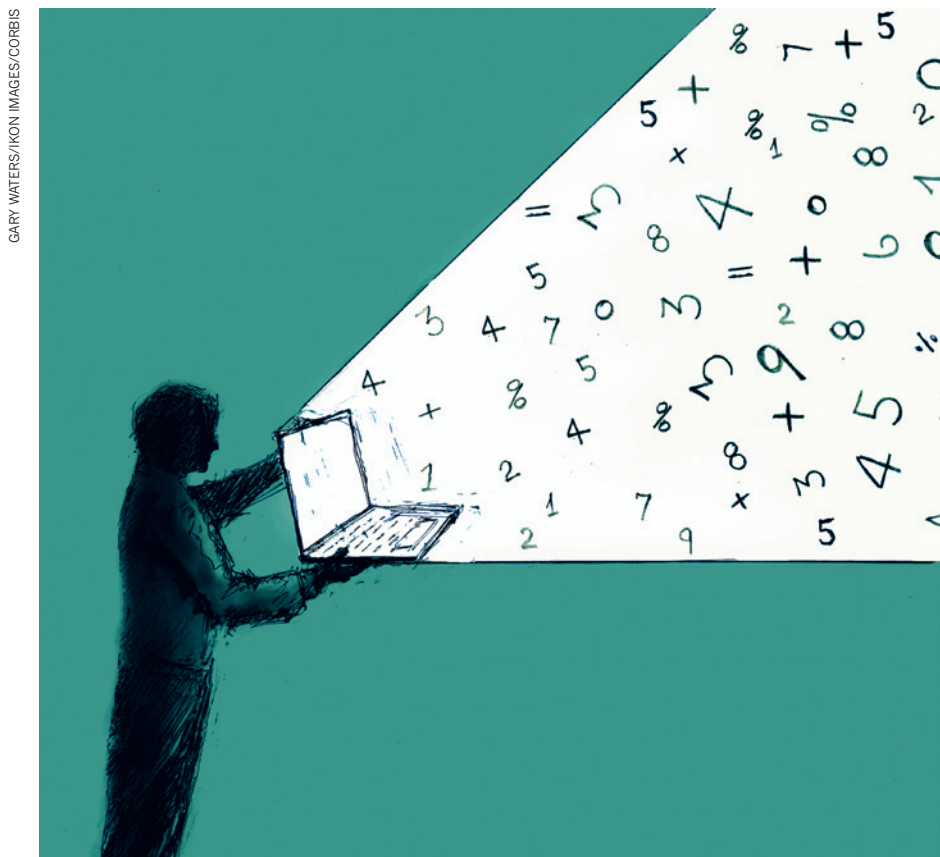The Cancer Genome Atlas Research Network

In the 'Results' section of this Article, the range in the sentence "The median follow-up of the cohort was 32 months (range, 1–19 months); 21% of the patients have recurred, and 11% have died." should have been 1–195 months. This error has been corrected in the HTML and PDF versions of the paper.

CAREERS

GARY WATERS/IKON IMAGES/CORBIS

DATA-SHARING

Everything on display

Researchers can get visibility and connections by putting their data online — if they go about it in the right way.

BY RICHARD VAN NOORDEN

Lizzie Wolkovich always felt she ought to make her research data freely available online. "The idea that data should be public has been in the background through my entire career," she says.

Yet in 2003–09, while she was working on her ecology PhD, there were few incentives for her to share. Sharing would not help her to get grants or publications, and although posting data online was not unheard of, few researchers actually did it, she says. Many preferred to hang on to their hard-won field data, sharing privately if they did so at all.

But after she earned her doctorate, Wolkovich overcame her hesitation, thanks to a combination of helpful colleagues, improved resources and a discernible shift in the research community's attitude. So in 2010, through an online data repository called the Knowledge Network for Biocomplexity, Wolkovich released her doctoral data set — the fruit of thousands of hours spent measuring the diversity of arthropods in 56 experimental soil plots she had set up in the arid scrubscape of southern California. Since then, she has publicized all the data that she has collected, including a meta-analysis of 50 other studies that she examined to see how factors such as rising temperatures affect the life cycles of plants. Wolkovich, now at the University of British Columbia in Vancouver, Canada, says that she herself had never objected to sharing her results — she had just not known how to do so. She likes the fact that her data are now easily accessible to other researchers and anyone else who is interested. "It saves me so much time," she says.

Wolkovich is one of a number of early-career researchers who are enthusiastically posting their work online. They are publishing what one online-repository founder calls small data — experimental results, data sets, papers, posters and other material from individual research groups — as opposed to the 'big data' spawned by large consortia, which usually employ specialists to plan their data storage and release. The many resources now available give researchers options for where and how to post their data, releasing potentially fruitful data sets that used to be locked up in unpublished paper files, buried in journal-article appendices or hidden away on scientists' hard drives.

OPENING UP

Open data-sharers are still in the minority in many fields. Although many researchers broadly agree that public access to raw data would accelerate science — because other scientists might be able to make advances not foreseen by the data's producers — most are reluctant to post the results of their own labours online (see *Nature* **461,** 160–163; 2009). When Wolkovich, for instance, went hunting for the data from the 50 studies in her meta-analysis, only 8 data sets were available online, and many of the researchers whom she e-mailed refused to share their work. Forced to extract data from tables or figures in publications, Wolkovich's team could conduct only limited analyses.

Some communities have agreed to share online — geneticists, for example, post DNA sequences at the GenBank repository, and astronomers are accustomed to accessing images of galaxies and stars from, say, the Sloan Digital Sky Survey, a telescope that has observed some 500 million objects — but these remain ▶

the exception, not the rule. Historically, scientists have objected to sharing for many reasons: it is a lot of work; until recently, good databases did not exist; grant funders were not pushing for sharing; it has been difficult to agree on standards for formatting data and the contextual information called metadata; and there is no agreed way to assign credit for data.

But the barriers are disappearing, in part because journals and funding agencies worldwide are encouraging scientists to make their data public. Last year, the Royal Society in London said in its report *Science as an Open Enterprise* that scientists need to "shift away from a research culture where data is viewed as a private preserve". Funding agencies note that data paid for with public money should be public information, and the scientific community is recognizing that data can now be shared digitally in ways that were not possible before. To match the growing demand, services are springing up to make it easier to publish research products online and enable other researchers to discover and cite them. There are so many, in fact, that choosing where and how to publish data sets and other supplementary material can be confusing (see 'Abundant options').

"Lots of people are getting into data-hosting, and I think it will be tricky to decide where to put your data," says Heather Piwowar, who studies data-sharing for the US National Evolutionary Synthesis Center in Durham, North Carolina.

SHARE AND SHARE ALIKE

Although exhortations to share data often concentrate on the moral advantages of sharing, the practice is not purely altruistic. Researchers who share get plenty of personal benefits, including more connections with colleagues, improved visibility and increased citations. The most successful sharers — those whose data are downloaded and cited the most often — get noticed, and their work gets used. For example, one of the most popular data sets on multidisciplinary repository Dryad is about wood density around the world; it has been downloaded 5,700 times. Co-author Amy Zanne, a biologist at George Washington University in Washington DC, thinks that users probably range from climate-change researchers wanting to estimate how much carbon is stored in biomass, to foresters looking for information on different grades of timber. "I would much prefer to have my data used by the maximum number of people to ask their own questions," she says. "It's important to allow readers and reviewers to see exactly how you arrive at your results. Publishing data and code allows your science to be reproducible."

Even people whose data are less popular can benefit, adds Piwowar. By making the effort to organize and label files so that others can understand them, scientists become more organized and better disciplined themselves, and can avoid confusion later on. "It is often very hard to find and understand your own work if you are looking at it years from now," says Piwowar. Scientists might be inclined to stuff their data into folders that can get lost and muddled — but if they store the files in an online repository, they are forced to curate and collate the data, she says.

The fear of being scooped is a powerful inhibitor. But scientists can put an embargo on their data, so that only they can see the work until they are ready to make it public. And data sets are becoming increasingly citable, bringing their authors formal recognition: data published in a data journal, on Dryad or on the repository figshare.com are given a digital object identifier (DOI) that can be referenced in other publications. (Figshare is owned by Digital Science, a sister company to Nature Publishing Group.)

Would-be sharers often worry that their data are too disordered or shoddy to release into the world. "I make my data available, and it can be a pain. I'm also scared and embarrassed about errors — most of us are, especially early-career scientists," says Piwowar. "We

don't yet have a culture of forgiveness around that, unlike in computer programming, where everyone knows there are bugs in code." She advises researchers to look into repositories to get a sense of the quality standard for experimental data. "It doesn't have to be perfect," she says. "It's probably less thorough than you think."

As sharing grows more common, scientists may worry less about posting data sets. "Ultimately, data will be so ubiquitous that we will no longer be in a world where researchers are so scared," says Carl Boettiger, an ecologist at the University of California, Santa Cruz, who keeps his entire laboratory notebook open online (see *Nature* **493,** 711; 2013). "At the end of the day, science is a social process. You will never get there hiding yourself and your work," he adds.

"Lots of people are getting into data-hosting, and I think it will be tricky to decide where to put your data."
Heather Piwowar

THE RIGHT PLACE

Depositing data on a personal website is unlikely to be the best way to get it reused and cited. For a start, the website may not be around in five years, says William Michener, director of e-science initiatives at the University of New Mexico in Albuquerque. Michener is principal investigator for a multinational programme called DataONE, which is funded by the US National Science Foundation and promotes best practices to scientists as part of its aim to make data more discoverable. Journal publishers back up their research papers with the help of non-profit archiving services such as Portico and CLOCKSS, which are financed by participating libraries and publishers, and which store material on a number of servers so that it will not disappear if a publisher goes bankrupt. Some data publishers have similar contingency plans, and Piwowar recommends looking into them. If no back-up plans are in place, she says, "it suggests they haven't prioritized well enough how to steward their data".

Just as important as sharing data publicly is making sure that other researchers can understand them. Susanna Assunta-Sansone, associate director of the Oxford e-Research Centre at the University of Oxford, UK, says that putting out data without noting what it means will ensure that "it's not really reusable". To avoid this, researchers must choose appropriate metadata: descriptions of the data's content and how they are arranged and set up. This type of curation is useful not just for human readers, but also for computer programmes that might be used to search

---

WHERE AND HOW
Abundant options

Online data repositories are proliferating: the searchable catalogue Databib lists 594 websites. Hundreds are specialists, devoted to particular kinds of data. But general-purpose repositories do exist: they include Dryad, which many scientists use to store the data underlying their publications; GitHub, which is usually used to host software code and to collaborate on developing it, but also hosts other data; European Commission repository ZENODO; and figshare.com, a general repository for posters, papers and data sets that welcomes negative results that would otherwise never be published.

Publishers have started to launch journals dedicated to data sets and descriptions of data, such as BioMed Central's *GigaScience*. Some scientists post data on social networks such as ResearchGate or Academia.edu.

Each discipline is evolving its own ways to structure data and metadata. In biology alone, biosharing.org lists some 530 standards, including MIAME (Minimum Information About a Microarray Experiment) and PDB (Protein Data Bank format). To avoid confusion, researchers should familiarize themselves with the best practices in their fields. **R.V.N.**

through or connect data sets. Intelligent searches often rely on whatever descriptive metadata researchers have attached to the data. The metadata are read by an application programming interface (API), a set of commands that computer programmes use to interact with data stores and pull information from them. Not all data repositories use APIs; those that do not may not be the best places to store or release information, because it could be hard for anyone to find.

Sites that are dedicated to hosting particular types of data, such as DNA sequences, usually tell submitters what format is appropriate. They may require data to be entered using an online form or following specific instructions. By contrast, generalist sites — such as institutional repositories, data journals or ventures similar to figshare.com — may have looser requirements. This has the potential to result in a blizzard of different formats and descriptive tags, which could make discovering and reusing data more difficult, so researchers should pay close attention to the norms in their fields.

Decisions about metadata standards should be made early in a research project, says Michener. DataONE has provided a primer on best practices, as has a tool called DataUp, run through the University of California Curation Center in Oakland to help researchers to create data packages that are good enough to put online. Other aspects of data-sharing to consider early on include the information's sensitivity and whether some parts must be stripped out to avoid, for example, identifying human study participants or the locations of endangered species. Researchers also need to be clear about whether they will allow their data sets to be used for any purpose, or whether they would like to limit reuse to, for example, non-commercial applications. One widely understood way of documenting reuse rights is by giving the data one of several different Creative Commons licences.

Ultimately, says Michener, early-career researchers need to pay attention to new and developing ways to share data, and to the standardized formats that are emerging to make data easier to search and discover. Those who do not, he says, should rethink why they are doing research. "I think we are just now reconnecting with what science is all about — not just creating new knowledge, but also sharing the information and data that underpins those discoveries." ∎

Richard Van Noorden *is a senior reporter at* Nature.

"At the end of the day, science is a social process. You will never get there hiding yourself and your work."

TURNING POINT
Kevin Gurney

Sustainability scientist Kevin Gurney has been studying climate change for 27 years. He has worked in academia, public policy, non-governmental organizations (NGOs) and think tanks, and is currently at Arizona State University in Tempe. He describes how he navigates the science–policy divide.

What convinced you to do a graduate degree?
As an undergraduate, I worked at the Lawrence Berkeley National Laboratory in California, taking spectroscopic measures of greenhouse gases. Working with wonderful mentors who were excited about the science was infectious. Later I did a master's in atmospheric science at the Massachusetts Institute of Technology in Cambridge and my focus shifted to chemistry and chlorofluorocarbons (CFCs) — greenhouse gases that also deplete Earth's ozone layer, and so have science and policy implications.

How did you become active in policy?
Regulation was ramping up to stop production of fully fluorinated CFCs, and industry was looking for alternatives. In 1986, I found that compounds called HCFCs, which contained less chlorine and thus caused less ozone depletion, still had the heat-trapping properties of CFCs. The policy implications were huge and there was so much misinformation. I was thinking, people need to know about this. I got more involved with policy at that point.

Why not go on immediately to pursue a PhD?
I wanted to work on the political implications first. In 1992, I started working with the Institute for Energy and Environmental Research in Takoma Park, Maryland. We sued the US Environmental Protection Agency to get it to regulate HCFCs, and we spread the word that HCFCs were not as environmentally friendly as manufacturers claimed. I also got involved in discussions on the Montreal Protocol, the treaty to regulate ozone-depleting chemicals. I realized how ineffectively science and policy interacted. I got a master's in public policy at the University of California, Berkeley, then a PhD in ecology at Colorado State University in Fort Collins. These days it is easier to get an interdisciplinary degree, but I tell my students that some degrees lack a rigorous science foundation. There is no substitute for a solid mathematics and physics background — it gives you credibility.

How did you move from CFCs to carbon?
I attended the negotiations in London and Copenhagen to amend the Montreal Protocol, laying out a plan to manage CFC phase-out.

Once the treaty was set, I began to see that rising carbon dioxide levels were an interesting problem. I maintained a personal network of contacts in NGOs, and many organizations were shifting to carbon dioxide and climate research for exactly the same reasons I was — it was quickly gaining traction. NGOs, including the US branch of the conservation group WWF in Washington DC, paid for me to go to Kyoto Protocol negotiations, and I worked pro bono as a science consultant. I told the NGOs I was not going to give anyone just a line they wanted to hear. My PhD adviser let me take vacation to attend negotiations every four months.

What is climate-change negotiation like?
It is the most intense, pressure-filled world you can imagine. I was very involved with language in the Kyoto Protocol about the missing carbon sink — the carbon dioxide absorbed on land, which is not fully understood — and how to account for it. I learned a lot about law during my policy degree, which made me effective in crossing the divide between policy and science. You don't have to dumb down; you have to learn how legislators and policy-makers view science.

You won a Faculty Early Career Development award from the US National Science Foundation in 2009. How are you using it?
I'm doing a risky thing and getting involved with citizen science to use Google Earth to identify power plants (see *Nature* http://doi.org/nb3; 2013). Normally I would be too worried that it would fail to use funding dollars. But we have thousands of people involved and are adding hundreds of power plants to an emissions database that is part of NASA's pilot carbon-monitoring system. It is of interest to climate scientists, social scientists and policy-makers. ∎

INTERVIEW BY VIRGINIA GEWIN

ONE MEAL A DAY

An unexpected guest.

BY MILO JAMES FOWLER

First thing Monday morning, Howard Schlange entered the living room of his one-bedroom apartment to find a well-dressed stranger sitting on the couch.

"Who are you?" Howard dropped back a step, his heart lurching against his chest.

"It is time for you to return," said the stranger, remaining seated. Only his head had turned, the hollow eyes in his pale face fixed upon Howard. "You have been here long enough."

Howard couldn't argue with that. Ever since construction on the freeway overpass had begun in earnest last month, the whole vibe of the neighbourhood had deteriorated. "You look like an undertaker."

"That does not change the situation."

"No, I guess not." Keeping a wary eye on the intruder, Howard edged to the bar that separated the living room from his kitchenette.

"Do not think to destroy me with a particle beam. It obviously did not work the last time."

"Obviously," Howard repeated, grasping his wallet and keys and holding them up. "I have an appointment."

"With destiny, yes."

"No, with a dentist." Howard licked his lips. His heart rate hadn't decelerated from high gear.

"Your grasp of this primitive language eludes me. How do you manage to hold that form?" A ripple coursed through the intruder's body as if there was a python coiled where his intestines should have been. "It is all I can do not to burst apart at the seams."

Howard blinked. "Listen, I don't have any cash —"

"You will not need *cash* — whatever it is — where we are going."

True enough; the insurance company would cover just about anything. "Wait. You're coming with me?"

"No. *You* are coming with *me*."

"That's kind of what I said." Howard stuffed the wallet into his back pocket and fiddled with the keys, jangling them against each other.

"What a dreadful sound!" The stranger's white, four-knuckled fingers flew to plug his ears. "Stop it at once!"

"Sorry." Howard dropped the keys into his pocket and glanced at the front door, which showed no signs of forced entry. Neither did the balcony's
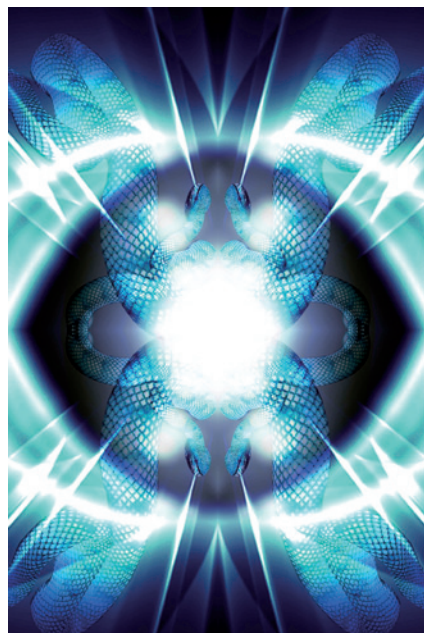
sliding door. Both were locked up, just the way he'd left them the night before. "How did you get in here exactly?"

"Much better." Mr Four Knuckles dropped his hands from his ears and stood suddenly, like a robot straightening itself. "Let us be on our way."

"Okay?" Howard moved towards the door.

"Where are you going?" the stranger snapped.

"The dentist. I told you —"

"There are far greater matters at stake, Prince Orionhart!"

"Orion-*who*?"

"Your father's kingdom is crumbling, Your Highness. The Crustaceoids lie at his very gates!" The impeccably dressed intruder came three steps forward, stomping his legs like stilts. "I realize we have had our differences in the past, that you never appreciated my meddling in your affairs, but you have to know that my only aim was to keep you safe from harm. Of course, as is the case with all youth, there comes a time when you desire to strike out on your own, leaving your doting caretakers behind, so I can understand why you attempted to kill me —"

"*Huh*?"

"Have you no memory of it at all?"

"It never happened — not with *me*, anyway." Howard tried to swallow but found a dry tongue in the way. "I think you've got me confused with somebody else. Mr ____?" He waited for the stranger to fill in the blank.

Instead, the fellow started muttering to himself: "I suppose it is possible that his brain chemistry could have undergone changes when he assumed the shape of one of these Earth creatures. I myself feel quite out of sorts. But it is highly unlikely that he would experience the sort of amnesia he appears to be exhibiting."

Howard almost smiled. "Oh, I get it. You think I'm an *alien*?" This was so cool all of a sudden!

The stranger scowled. "The oxygen must be affecting you —"

"You think I'm an alien *prince*? Holy cow!" Howard couldn't wipe the grin from his face. The weird visitor's skin began to contort as if underneath it was a nest of spider eggs that were all hatching at once. "You want to take me to your mothership?"

"I am not acquainted with the female, but if you are referring to the Cosmic Conveyor —"

Howard let out a whoop. "This is freakin' *awesome*!" All thoughts of his appointment vanished from his mind. He stood to attention, doing his best to contain himself. "All righty then. Take me to your leader."

The stranger blinked. "*You* are my leader."

"Right. So, take me back to wherever. Out there." He gestured vaguely at the ceiling.

The stranger's eyes squinted oddly as if the lower lids were attempting to devour his eyeballs whole. "I am beginning to have my doubts."

"I'm Prince Orionhead, ready to return to my homeworld. Beam me up!"

The stranger whipped a chrome pen from his breast pocket and pressed the tip with one of his long fingers. A white light enveloped him. "I believe I have made a mistake …"

"No mistake!" Howard charged into the light … And ended up sprawled out on his couch, very much alone.

"Hey, what about me?" he shouted at the ceiling.

In the silence that followed, he heard only his stomach growling. How long had it been since he'd eaten last? After yesterday's mail arrived?

He looked down at his T-shirt and smirked at what appeared to be a pair of boa constrictors squirming inside his protruding belly. Shrugging, he headed out for his appointment.

Mailman had been a fine delicacy, but he couldn't wait to sample *dentist*. ∎

Milo James Fowler *is an English teacher by day and a speculative fictioneer by night. Find him at www.milo-inmediasres.blogspot.co.uk.*